# This Page Is Inserted by IFW Operations and is not a part of the Official Record

## **BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

## IMAGES ARE BEST AVAILABLE COPY.

As rescanning documents will not correct images, please do not report the images to the Image Problem Mailbox.

THIS PAGE BLANK (USPTO)

#### **PCT**

#### WORLD INTELLECTUAL PROPERTY ORGANIZATION International Bureau



#### INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classificati n 6: C12Q 1/68, G01N 33/53, C12P 19/34,

(11) International Publication Number:

WO 96/25519

C12N 5/10, 1/21, C07K 5/00, 14/00, 16/00, C07H 21/02, 21/04

(43) International Publication Date:

22 August 1996 (22.08.96)

(21) International Application Number:

PCT/US96/01938

(22) International Filing Date:

15 February 1996 (15.02.96)

(30) Pri rity Data:

08/390,878

US 17 February 1995 (17.02.95)

(71) Applicant: PATHOGENESIS CORPORATION [US/US]; Suite 150, 201 Elliott Avenue West, Seattle, WA 98119 (US).

(72) Inventors: STOVER, Charles, Kendall; 7640 81st Place S.E., Mercer Island, WA 98040 (US). MAHAIRAS, Gregory, G.; 3312 39th West, Seattle, WA 98199 (US).

(74) Agents: HUNTER, Tom et al.; Townsend and Townsend and Crew, Steuart Street Tower, One Market, San Francisco, CA 94105-1492 (US).

(81) Designated States: AL, AM, AT, AU, AZ, BB, BG, BR, BY, CA, CH, CN, CZ, DE, DK, EE, ES, FI, GB, GE, HU, IS, JP, KE, KG, KP, KR, KZ, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, TJ, TM, TR, TT, UA, UG, UZ, VN, ARIPO patent (KE, LS, MW, SD, SZ, UG), Eurasian patent (AZ, BY, KG, KZ, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).

#### **Published**

With international search report.

(54) Title: VIRULENCE-ATTENUATING GENETIC DELETIONS

#### (57) Abstract

The present invention provides specific genetic deletions that result in an avirulent phenotype of a mycobacterium. These deletions may be used as phenotypic markers of providing a means for distinguishing between disease-producing and non-disease producing mycobacteria.

#### FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AM	Armenia	GB	United Kingdom	MW	Malawi
AT	Austria	GE	Georgia	MX	Mexico
ΑU	Australia	GN	Guinea	NE	Niger
BB	Barbados	GR	Greece	NL	Netherlands
BE	Belgium	HU	Hungary	NO	Norway
BF	Burkina Faso	IE	Ireland	NZ	New Zealand
BG	Bulgaria	ſΤ	Italy	PL	Poland
BJ	Benin	JP	Japan	PT	Portugal
BR	Brazil	KE	Kenya	RO	Romania
BY	Belarus	KG	Kyrgystan	RU	Russian Federation
CA	Canada	KP	Democratic People's Republic	SD	Sudan
CF	Central African Republic		of Korea	SE	Sweden
CG	Congo	KR	Republic of Korea	SG	Singapore
CH	Switzerland	KZ	Kazakhstan	Si	Slovenia
CI	Côte d'Ivoire	LI	Liechtenstein	SK	Slovakia
CM	Cameroon	LK	Sri Lanka	SN	Senegal
CN	China	LR	Liberia	SZ	Swaziland
CS	Czechoslovakia	LT	Lithuania	TD	Chad
CZ	Czech Republic	LU	Luxembourg	TG	Togo
DE	Germany	LV	Larvia	TJ	Tajikistan
DK	Denmark	MC	Monaco	TT	Trinidad and Tobago
EE	Estonia	MD	Republic of Moldova	UA	Ukraine
ES	Spain	MG	Madagascar	UG	Uganda
FL	Finland	ML	Mali	US	United States of America
FR	France	MN	Mongolia ·	UZ	Uzbekistan
GA	Gabon	MR	Mauritania	VN	Viet Nam

10

15

20

25

30

# VIRULENCE-ATTENUATING GENETIC DELETIONS

#### BACKGROUND OF THE INVENTION

Mycobacterium nuberculosis (MTB) infects over ten million people each year and kills over three million, making it the infectious agent causing the greatest mortality worldwide. In an effort to combat Mycobacterium nuberculosis, vaccination programs using a viable attenuated strain of Mycobacterium bovis called bacille Calmette-Guérin (BCG) have been established in more than 120 countries over the course of the last 5 decades. Although widely used and considered safe enough to administer to infants, the BCG vaccine is controversial for two principle reasons: 1) Efficacy for BCG vaccines against tuberculosis has varied from 0-85% in different clinical trials; and 2) Immunization with BCG sensitizes vaccinees to the tubercular antigens used in the tuberculin skin test, confounding attempts to discriminate between BCG immunization and TB infection. For these two reasons, especially the latter, BCG is not used in the United States where surveillance with the tuberculin test is preferred.

The original Pasteur BCG strain was developed by multiple (230 times) serial passages in liquid culture. BCG has never been shown to revert to virulence in animals indicating that the attenuating mutations in BCG are stable deletions and/or multiple mutations which cannot revert. However, the mutations which arose during serial passage of the original BCG strain have never been identified. Moreover, recent efforts to genetically complement BCG virulence with genomic libraries of virulent tubercle bacilli have also been unsuccessful again suggesting that multiple unlinked mutations are responsible for the attenuation of BCG virulence. The antigenicity of BCG and the characteristics leading to its avirulence are thus poorly understood.

#### SUMMARY OF THE INVENTION

The present invention provides specific genetic deletions that account for the avirulent phenotype of the bacille Calmette-Guérin (BCG) strain of *Mycobacterium bovis*. These deletions may be used as phenotypic markers of providing a means for distinguishing between disease-producing and non-disease producing mycobacteria.

10

15

20

25

In a preferred embodiment, this invention provides for nucleic acid sequences that are markers for avirulent or virulent mycobacteria. The sequences uniquely characterize the presence or absence of deletions that result in an avirulent phenotype. More specifically the sequence are either deletion junction sequence or deletion sequences or subsequences within deletion junction sequences or deletion sequences. Thus, this invention provides for a marker for an avirulent mycobacterium comprising a first nucleic acid that hybridizes under stringent conditions with a second nucleic acid or a complement of the second nucleic acid where the second nucleic acid or its complement includes BCGa1a, BCGa1b, BCGa2a, BCGa2b, BCGa3a, BCGa3b, BCGa1ab, BCGa2ab, BCGa3ab, BCGa1, BCGa2, and BCGa3. In a particularly preferred embodiment, the marker specifically hybridizes under stringent conditions to a nucleic acid from BCG, but not to a nucleic acid from Mycobacterium tuberculosis of Mycobacterium bovis. or alternatively, the marker specifically hybridizes under stringent conditions to a nucleic acid from Mycobacterium tuberculosis or Mycobacterium bovis, but not to a nucleic acid from BCG. The marker may be the full length BCGala, BCGalb, BCGa2a, BCGa2b, BCGa3a, BCGa3b, BCGalab, BCGa2ab, BCGa3ab, BCGa1, BCGa2, and BCGa3 or a subsequence within any of these regions. The marker may also include a nucleic acid having at least 80%, preferably 90%, more preferably 95%, and most preferably 98% percent sequence identity with BCGA1a. BCGalb, BCGa2a, BCGa2b, BCGa3a, BCGa3b, BCGa1ab, BCGa2ab, BCGa3ab, BCGa1, BCGa2, or BCGa3. The marker may also include a sequence selected from an open reading frame of a the deletion sequences BCGa1, BCGa2, BCGa3. Suitable open reading frames are indicated in Figures 4, 5, and 6.

The above described marker may be a probe. The probe may be labeled by a number of means including, but not limited to radioactive, fluorescent, enzymatic, and colorimetric labels.

In another embodiment, this invention provides for polypeptides encoded by a subsequence of the BCGA1, BCGA2, or BCGA3 deletions. In particular, the subsequence may be selected from an open reading frame (ORF) present in one of these deletion sequences. This invention also provides for monoclonal or polyclonal antibodies that

10

15

20

25

30

specifically bind polypeptides encoded by one or more subsequences of the BCGa1, BCGa2, or BCGa3 deletions.

In still another embodiment, this invention provides for a recombinant cell comprising a first nucleic acid that hybridizes under stringent conditions with a second nucleic acid or a complement of the second nucleic acid where the second nucleic acid or its complement is BCGala, BCGalb, BCGala, BCGala

In still yet another embodiment, this invention provides a method of distinguishing between an attenuated and a virulent mycobacterium. The method involves detecting the presence or absence of a first nucleic acid that hybridizes under stringent conditions with a second nucleic acid or a complement of the second nucleic acid where the second nucleic acid or its complement is BCGAla, BCGAlb, BCGA2a, BCGA2b, BCGA3a, BCGA3b, BCGAlab, BCGA2ab, BCGA3ab, BCGA1a, BCGA1b, BCGA2, or BCGA3. The first nucleic acid may include any of the markers described above. A particularly preferred marker is one that specifically hybridizes under stringent conditions to a nucleic acid from BCG, but not to a nucleic acid from Mycobacterium tuberculosis or Mycobacterium bovis, or alternatively, that specifically hybridizes under stringent conditions to a nucleic acid from Mycobacterium tuberculosis or Mycobacterium bovis, but not to a nucleic acid from BCG. The method may involve amplifying either the first nucleic acid by any of a number of methods including, for example, polymerase chain reaction. The detection may involve detecting the first nucleic acid, for example, as in a Southern blot, or alternatively, detecting a polypeptide encoded by the first nucleic acid. More specifically, the polypeptide may be

a encoded by an open reading frame (ORF) selected from BCGa1, BCGa2, or BCGa3. The polypeptide may be visualized by a number of means well known to those of skill in the art including antibody hybridization such as direct or indirect binding of labeled antibody.

5

10

15

This invention additionally provides a method for determining whether an attenuated or a virulent Mycobacterium is present in a sample. This method involves providing a first nucleic acid that hybridizes under stringent conditions with a second nucleic acid or a complement of the second nucleic acid where the second nucleic acid or its complement is BCGala, BCGalb, BCGa2a, BCGa2b, BCGa3a, BCGa3b, BCGalab, BCGa2ab, BCGa3ab, BCGa1, BCGa2, or BCGa3; and hybridizing the first nucleic acid to the biological sample. The first nucleic acid may include any of the markers described above. A particularly preferred marker is one that specifically hybridizes under stringent conditions to a nucleic acid from BCG, but not to a nucleic acid from Mycobacterium suberculosis or Mycobacterium bovis, or alternatively, that specifically hybridizes under stringent conditions to a nucleic acid from Mycobacterium tuberculosis or Mycobacterium bovis, but not to a nucleic acid from BCG. The method may involve amplifying either the first nucleic acid by any of a number of methods including, for example, polymerase chain reaction. The detection may involve detecting the first nucleic acid, for example, as in a Southern blot, or alternatively, detecting a polypeptide encoded by the first nucleic acid. More specifically, the polypeptide may be a encoded by an open reading frame (ORF) selected from BCGa1, BCGa2, or BCGa3. The method may also include detecting the hybridized first nucleic acid. This may involve direct detection of a label or additionally involve an amplification step and subsequent detection of the amplified product.

25

30

20

Finally, this invention provides a method of producing an attenuated-virulence mycobacterium. This method involves deleting from the genomic DNA of a virulent mycobacterium a first nucleic acid that specifically hybridizes under stringent conditions with a second nucleic acid or a complement of said second nucleic acid where said second nucleic acid or complement of said second nucleic acid is selected from the group consisting of BCGA1, BCGA2, and BCGA3. The first nucleic acid may be BCGA1, BCGA2, or BCGA3, or alternatively, it may be a promoter, other control element or an open reading frame from BCGA1, BCGA2, or BCGA3.

#### **Definitions**

5

10

15

20

25

30

Although any methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present invention, the preferred methods and materials are described. For purposes of the present invention, the following terms are defined below.

The phrase "specifically detect" as used herein refers to the process of determining that a particular subsequence is present in a DNA sample. A DNA sequence may be specifically detected through a number of means known to those of skill in the art. These would include, but are not limited to amplification of the particular target sequence through polymerase chain reaction or ligase chain reaction, hybridization of the sequence to a labeled probe, and binding by labelled ligands or monoclonal antibodies. For a discussion of various means of detection of specific nucleic acid sequences see Perbal, B. A Practical Guide to Molecular Cloning, 2nd Ed. John Wiley & Sons, N.Y. (1988) which is incorporated herein by reference.

The phrase "select subsequence" is used herein to refer to a particular DNA subsequence that is of interest. It is often a predetermined or known sequence of nucleic acid bases. A select subsequence is typically chosen because of a unique sequence identity. Typically a select subsequence is targeted for DNA amplification and often is useful as a specific marker for the presence of a particular gene or a deletion of a particular nucleic acid sequence.

The term "oligonucleotide" refers to a molecule comprised of two or more deoxyribonucleotides or ribonucleotides. Oligonucleotides may include, but are not limited to, primers, probes, nucleic acid fragments to be detected, and nucleic acid controls. Oligonucleotides include naturally occurring nucleotides, chemically modified naturally occurring nucleotides and synthetic nucleotides. The exact size of an oligonucleotide depends on many factors and the ultimate function or use of the oligonucleotide.

The term "primer" refers to an oligonucleotide, whether natural or synthetic, capable of acting as a point of initiation of DNA synthesis under conditions in which synthesis of a primer extension product complementary to a nucleic acid strand is induced, i.e., in the presence of four different nucleoside triphosphates and an agent for polymerization (i.e., DNA polymerase or reverse transcriptase) in an appropriate buffer and at a suitable temperature. A primer is preferably a single-stranded oligodeoxyribonucleotide.

10

15

20

25

30

The appropriate length of a primer depends on the intended use of the primer but typically ranges from 15 to 25 nucleotides. Short primer molecules generally require cooler temperatures to form sufficiently stable hybrid complexes with the template. A primer need not reflect the exact sequence of the template but must be sufficiently complementary to hybridize with a template.

The phrase "PCR primers competent to amplify" as used herein refers to a pair of PCR primers whose sequences are complementary to DNA subsequences immediately flanking the DNA subsequence (target sequence) which it is desired to amplify. The primers are chosen to bind specifically those particular flanking subsequences and no other sequences present in the sample. The PCR primers are thus preferably chosen to amplify the unique target sequence and no other. Alternatively, the PCR primers may be selected to bind to sequences other than the target sequence where the amplification products can be subsequently distinguished (e.g. where the desired amplified sequence is different in size than other amplified sequences).

"Amplifying" or "amplification", which typically refer to an "exponential" increase in target nucleic acid, are used herein to describe both linear and exponential increases in the number of a select target sequence of nucleic acid.

The term "antisense orientation" refers to the orientation of nucleic acid sequence from a structural gene that is inserted in an expression cassette in an inverted manner with respect to its naturally occurring orientation. When the sequence is double stranded, the strand that is the template strand in the naturally occurring orientation becomes the coding strand, and vice versa.

The term "deletion" refers to a region of a nucleic acid which is not present in an organism, but which is present in another related organism. In the context of mycobacteria, a deletion refers, e.g., to a region of nucleic acid which is not present in one strain of mycobacteria, but which is present in another related strain. For instance, an avirulent mycobacterial strain can have a deletion in its genome relative to the genome of a related virulent mycobacterial strain.

The term "deletion junction" refers to the region of a nucleic acid spanning the insertion point of a deletion. Thus, where a region of a nucleic acid sequence is deleted (i.e. a deletion is present), the deletion junction spans the nucleotides that are immediately adjacent to the deletion. Conversely, where a region of a nucleic acid sequence is not

deleted (i.e. the deletion is absent), two deletion junctions are present, each spanning respectively one end of the deletion sequence and its flanking sequence.

The following terms are used to describe the sequence relationships between two or more polynucleotides: "reference sequence", "comparison window", "sequence identity", "percentage of sequence identity", and "substantial identity". A "reference sequence" is a defined sequence used as a basis for a sequence comparison; a reference sequence may be a subset of a larger sequence, for example, as a segment of a full-length cDNA or gene sequence given in a sequence listing, such as a polynucleotide sequence of Figures 1, 2, or 3, or may comprise a complete cDNA or gene sequence.

10

15

5

Generally, a reference sequence is at least 10 nucleotides in length, frequently at least 20 to 25 nucleotides in length, and often at least 50 nucleotides in length. Sequence comparisons between two (or more) polynucleotides are typically performed by comparing sequences of the two polynucleotides over a "comparison window" to identify and compare local regions of sequence similarity. A "comparison window", as used herein, refers to a segment of at least 10 contiguous nucleotide positions wherein a polynucleotide sequence may be compared to a reference sequence of at least 10 contiguous nucleotides and wherein the portion of the polynucleotide sequence in the comparison window may comprise additions or deletions (i.e., gaps) of 20 percent or less as compared to the reference sequence (which does not comprise additions or deletions) for optimal alignment of the two sequences.

20

25

Optimal alignment of sequences for aligning a comparison window may be conducted by the local homology algorithm of Smith and Waterman Adv. Appl. Math. 2: 482 (1981), by the homology alignment algorithm of Needleman and Wunsch J. Mol. Biol. 48: 443 (1970); by the search for similarity method of Pearson and Lipman Proc. Natl. Acad. Sci. (USA) 85: 2444 (1988), or by computerized implementations of these algorithms (GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package Release 7.0, Genetics Computer Group, 575 Science Dr., Madison, WI), or by inspection, and the best alignment (i.e., resulting in the highest percentage of sequence similarity over the comparison window) generated by the various methods is selected.

30

The term "sequence identity" means that two polynucleotide sequences are identical (i.e., on a nucleotide-by-nucleotide basis) over the window of comparison. The term "percentage of sequence identity" is calculated by comparing two optimally aligned

10

15

20

25

30

sequences over the window of comparison, determining the number of positions at which the identical nucleic acid base (e.g., A, T, C, G, U, or I) occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the window of comparison (i.e., the window size), and multiplying the result by 100 to yield the percentage of sequence identity. The term "identical" in the context of two nucleic acid or polypeptide sequences refers to the residues in the two sequences which are the same when aligned for maximum correspondence.

The terms "isolated" or "biologically pure" refer to material which is substantially or essentially free from components which normally accompany it as found in its native state. The isolated nucleic acid probes of this invention do not contain materials normally associated with their in situ environment, in particular nuclear, cytosolic or membrane associated proteins or nucleic acids other than those nucleic acids intended to comprise the nucleic acid probe itself.

The term "marker" refers to a characteristic which distinguishes one class of cells or compositions from a second class of cells or compositions. For instance, the deletions and deletion junctions described herein can be used to distinguish between strains (e.g., virulent and avirulent strains) of mycobacteria. While markers are indicators of associated features or properties, as used herein, markers may also be used for purposes other than indicating the associated feature or property. Thus, for example, a nucleic acid marker of virulence identifies a particular nucleic acid which may be used in a variety of contexts other than simply indicating virulence.

The term "nucleic acid" refers to a deoxyribonucleotide or ribonucleotide polymer in either single- or double-stranded form, and unless otherwise limited, encompassing known analogues of natural nucleotides that can function in a similar manner as naturally occurring nucleotides. Unless otherwise indicated, a particular nucleic acid sequence includes the complementary sequence thereof.

The term "operably linked" refers to functional linkage between a promoter and a second sequence, wherein the promoter sequence initiates transcription of RNA corresponding to the second sequence.

The term "peptide" or "polypeptide" refers to an amino acid polymer which is encoded by a nucleic acid. The peptide or polypeptide may include naturally occurring or modified amino acids.

٠. ٨

13

. .

5

10

15

20

25

30

The terms "probe" or "nucleic acid probe" refer to a molecule that binds to a specific sequence or subsequence of a nucleic acid. A probe is preferably a nucleic acid which binds through complementary base pairing to the full sequence or to a subsequence of a target nucleic acid. It will be understood by one of skill in the art that probes may bind target sequences lacking complete complementarily with the probe sequence depending upon the stringency of the hybridization conditions. The probes are preferably directly labelled as with isotopes, chromophores, lumiphores, chromogens, or indirectly labelled such with, e.g., biotin to which a streptavidin complex may later bind. By assaying for the presence or absence of the probe, one can detect the presence or absence of the selected sequence or subsequence.

The term "labeled nucleic acid probe" refers to a nucleic acid probe that is bound, either covalently, through a linker, or through ionic, van der Waals or hydrogen "bonds" to a label such that the presence of the probe may be detected by detecting the presence of the label bound to the probe.

The term "recombinant" when used with reference to a cell indicates that the cell replicates or expresses a nucleic acid, or expresses a peptide or protein encoded by DNA whose origin is exogenous to the cell. Recombinant cells can express genes that are not found within the native (non-recombinant) form of the cell. Recombinant cells can also express genes found in the native form of the cell wherein the genes are re-introduced into the cell by artificial means.

The term "sample" refers to a material with which bacteria may be associated. Frequently the sample will be a "clinical sample" which is a sample derived from a patient. Such samples include, but are not limited to, sputum, blood, blood cells (e.g., white cells), tissue or fine needle biopsy samples, urine, peritoneal fluid, and pleural fluid, or cells therefrom. It will be recognized that the term "sample" also includes supernatant from eukaryotic cell cultures (which may contain free bacteria), cells from cell or tissue culture, and other media in which it may be desirable to detect mycobacteria (e.g., food and water).

The term "subsequence" in the context of a particular nucleic acid sequence refers to a region of the nucleic acid equal to or smaller than the specified nucleic acid.

The term "substantial identity" or "substantial similarity" indicates that a nucleic acid or polypeptide comprises a sequence that has at least 90% sequence identity to a reference sequence, or preferably 95%, or more preferably 98% sequence identity to the

10

15

20

25

30

reference sequence, over a comparison window of at least about 10 to about 100 nucleotides or amino acid residues. An indication that two polypeptide sequences are substantially identical is that one protein is immunologically reactive with antibodies raised against the second protein. An indication that two nucleic acid sequences are substantially identical is that the polypeptides which the first nucleic acids encodes is immunologically cross reactive with the polypeptide encoded by the second nucleic acid.

Another indication that two nucleic acid sequences are substantially identical is that the two molecules hybridize to each other under stringent conditions. Stringent conditions are sequence-dependent and will be different with different environmental parameters. Generally, stringent conditions are selected to be about 5°C to 20°C lower than the thermal melting point (T<sub>m</sub>) for the specific sequence at a defined ionic strength and pH. The T<sub>m</sub> is the temperature (under defined ionic strength and pH) at which 50% of the target sequence hybridizes to a perfectly matched probe. Typically, stringent conditions will be those in which the salt concentration is at least about 0.2 molar at pH 7 and the temperature is at least about 60°C.

The term "uninterrupted reading frame" or "open reading frame" refers to a DNA sequence (e.g., cDNA) lacking a stop codon or other intervening, untranslated sequence. An intact open reading frame refers to a full length uninterrupted reading frame or minor variations thereof.

The term "virulent" in the context of mycobacteria refers to a bacterium or strain of bacteria that replicates within a host cell or animal at a rate that is detrimental to the cell or animal within its host range. More particularly virulent mycobacteria persist longer in a host than avirulent mycobacteria. Virulent mycobacteria are typically disease producing and infection leads to various disease states including fulminant disease in the lung, disseminated systemic milliary tuberculosis, tuberculosis meningitis, and tuberculosis abscesses of various tissues. Infection by virulent mycobacteria often results in death of the host organism. Typically, infection of guinea pigs is used as an assay for mycobacterial virulence. In contrast, the term "avirulent" refers to a bacterium or strain of bacteria that either does not replicate within a host cell or animal within its host range, or replicates at a rate that is not significantly detrimental to the cell or animal.

The term BCG-like avirulence, as used herein refers to an attenuated virulence brought about by one of the deletions of the present invention.

10

15

20

25

30

### BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 shows the complete sequence listing of the BCG deletion region 1 including flanking sequences. The deletion, designated BCGa1, is located between nucleotide 2327 and nucleotide 11126.

Figure 2 shows the complete sequence listing of the BCG deletion region 2 including flanking sequences. The deletion, designated BCGA2, is located between nucleotide 3382 and nucleotide 14071.

Figure 3 shows the complete sequence listing of the BCG deletion region 3 including flanking sequences. The deletion, designated BCG A3, is located between nucleotide 1406 and nucleotide 10673. "N" represents "A", "C", "G", or "T".

Figure 4 shows a map of the deletion sequence BCGa1. This map identifies the various open reading frames (ORFs) and indicates their location within the deletion sequence. Ribozome binding sites and homologies to the predicted encoded proteins are shown.

Figure 5 shows a map of the deletion sequence BCGA2. This map identifies the various open reading frames (ORFs) and indicates their location within the deletion sequence. Ribozomal binding sites and homologies to the predicted encoded proteins are shown.

Figure 6 shows a map of the deletion sequence BCGA3. This map identifies the various open reading frames (ORFs) and indicates their location within the deletion sequence. Ribozome binding sites and homologies to the predicted encoded proteins are shown. The sequence of a small region, estimated to be much less than 200 bp and located close to 9400 bp in Figure 3, remains to be determined. Therefore, the base pair coordinates given in the region 3 map 3' to the 9kb marker are approximations. The precise sequence determination of this region is likely to effect the length of open reading frames 3H and 3L.

Figure 7 illustrates the deletion junction regions of BCGa1, BCGa2, and BCGa3. The "terminal" deletion junction regions formed by the flanking sequences and the terminal regions of the deletion sequences are identified as BCGa1a, BCGa1b, BCGa2a, BCGa2b, and BCGa3a, and BCGa3b. When the deletion is present (the deletion sequences

10

15

20

25

30

are missing) the respective "a" and "b" sequences will be juxtaposed, thereby forming deletion "spanning" junction sequences designated BCG $\triangle$ 1ab, BCG $\triangle$ 2ab, and BCG $\triangle$ 3ab, respectively.

Figure 8 shows EcoRI and BamHI restricted chromosomal DNAs from Mycobacterium bovis, BCG Connaught, and Mycobacterium tuberculosis strains H37Ra, H37Rv, and Erdman probed with <sup>32</sup>P labeled BCG subtracted probe.

#### DETAILED DESCRIPTION

This invention reflects the discovery of genetic deletions in mycobacteria that result in an avirulent genotype such as is exhibited by the bacille Calmette-Guérin (BCG) mycobacterium. The original Pasteur bacille Calmette-Guérin (BCG) strain was developed by multiple (230 times) serial passages in liquid culture. BCG has never been shown to revert to virulence in animals indicating that the attenuating mutations in BCG are stable deletions and/or multiple mutations that cannot revert. The mutations that arose during serial passage of the original BCG strain were not previously known. Recent efforts to genetically complement BCG virulence with genomic libraries of virulent tubercle bacilli were unsuccessful, again suggesting that multiple unlinked mutations are responsible for the attenuation of BCG virulence.

The genetic deletions leading to the avirulent phenotype of BCG were identified by genomic subtractions between Connaught strain of BCG and MBV/MTB. The subtracted probe resulting from the genomic subtraction between BCG and the H37 Rv strain of M. tuberculosis was subsequently used to identify and clone three regions from a cosmid library of Mycobacterium bovis genomic DNA. Southern blot mapping and DNA sequence comparisons between BCG and M. bovis showed that three regions, designated regions 1-3, contained DNA segments of approximately 9 kb, 11 kb and 9 kb respectively, which are deleted in the Connaught strain of BCG. Precise deletion junctions were identified for each region by comparisons of BCG and corresponding virulent MBV sequences. The respective deletions, designated BCGA1, BCGA2 and BCGA3 are illustrated in Figures 1-3.

One of skill in the art will appreciate that the deletions encompassed by BCGa1, BCGa2 and BCGa3 may be utilized in a variety of contexts. For example, the deletions may be utilized to distinguish between avirulent and virulent strains of

10

15

20

25

30

mycobacteria thereby providing early detection of patients at risk for tuberculosis. This is of particular importance where mycobacteria are identified in a sample from a patient that has been previously vaccinated with BCG. In this context it may be critical to determine whether mycobacteria identified in a biological sample from such a patient are pathogenic.

In another embodiment, the preparation of mycobacteria containing the deletions of the present invention may provide superior vaccines to BCG which has long been known to have marginal efficacy. Thus, for example, a *Mycobacterium tuberculosis* may contain a full BCGal deletion or a smaller deletion within BCGal (e.g. one or more open reading frames) rendering it avirulent. An avirulent MTB will provide a more efficient vaccine because it is antigenically more similar to MTB than is BCG. Moreover, an MTB rendered avirulent by the production of smaller deletions within the deletion regions identified in this invention will present more antigenic determinants.

Since the loss of virulence is due to the loss of gene products expressed by the nucleic acid sequences comprising the deletion regions, the BCGA1, BCGA2 and BCGA3 deletion sequences and proteins encoded within these deletion sequences provide suitable targets for drug screening. Thus, the use of deleted sequences as targets to screen for drugs that inhibit or interfere with transcription, translation, or post-translational processing of proteins encoded by the deletion sequences, or with the deletion encoded polypeptides themselves, provides an assay for anti-mycobacterial agents. In particular, the use of reporter genes such as firefly luciferase (FFlux), \( \beta\)-galactosidase (BGal), and the like, under the control of promoters present in the deletion sequence provide a rapid assay for drugs regulating activity originating in this region. Conversely, since the protein products of the deletion sequences are presumably expressed in virulent mycobacterial species, proteins expressed by deletion sequences may make good antigens for antimycobacterial vaccines.

Finally, as the viability of BCG demonstrates, deletion regions BCGa1, BCGa2 and BCGa3 are not required for mycobacterial growth and reproduction. Thus, these deletion regions provide good insertion points for the expression of heterologous DNA. The heterologous DNA sequences may be under the control of endogenous inducible or constitutive promoters typically found in the deletion sequences, or alternatively, they may be under the control of introduced promoters, either constitutive or inducible, exogenous to mycobacteria.

10

15

20

25

30

#### I. Detection of Deletions

As indicated above, the deletions identified in the present invention provide useful markers for the identification of an avirulent (or conversely a virulent) mycobacterial phenotype. Specifically, determination of avirulence simply requires the detection of the presence or absence of the deletion (either BCGa1, BCGa2, or BCGa3, or deletions within these regions). Where the deletion is present in the bacterial DNA, the bacterium expresses a BCG-like avirulent phenotype. Conversely, where the deletion is absent in the bacterial DNA, the bacterium does not express a BCG-like avirulence. While this may indicate that the bacterium is virulent, one of skill will appreciate that the bacterium may still be avirulent due to the presence of other mutations or deletions. Nevertheless, screening for the presence of the deletion provides a means of detecting a BCG-like avirulent mycobacterium.

Means of detecting deletions are well known to those of skill in the art. Generally, the deletions may be detected either by detecting the presence or absence of deletion junctions, or, alternatively, by detecting the presence or absence of the sequences contained within the deletion (deletion sequences). Where a nucleic acid sequence is deleted (i.e., a deletion is present), the sequences that previously flanked the deleted sequence are juxtaposed, thereby forming a new deletion junction that spans the deletion. Detection of the presence of such a "spanning" deletion junction indicates the presence of the deletion and thus the avirulent phenotype.

Conversely, where the nucleic acid sequence is not deleted (the deletion is not present) the spanning junction sequence will be absent (See, e.g. Figure 7). The "terminal" deletion junction sequences flanking each endpoint of the deletion region are present and detection of these terminal deletion junctions indicates the absence of a deletion. Spanning deletion junction regions and terminal deletion junctions suitable for detecting the deletions of the present invention are illustrated in Figure 7 and in Table 1.

Table 1. Nucleic acid sequences comprising deletion junctions. The symbol "|" indicates the insertion point of the deletion sequence. Deletion sequence bases are represented in lower case letters.

Junction	Nucleotide Sequence	Seq.
BCG∆la	CTGGTCGACGATTGGCACAT   gcagccgtgggtgccgccgg	1

BCG△1b	gigiciticateggeticcae   CCAGCCGCCCGGATCCAGCA	2
BCG <sub>4</sub> 2a	CAACTCCACGGCGACCACCC   gcgcccccgctcgcactaga	3
BCG△2b	gcccacccggtcgagcaccc   CGATGATCTTCTGTTTGACC	4
BCG <sub>4</sub> 3a	CACCTCGACCACGGCCAACC   gtggacctgtgagatacact	5
BCG <sub>4</sub> 3b	tcagcagtccacggccaacc   CCGCACCAACACCTTCCACC	6
BCG△lab	CTGGTCGACGATTGGCACAT CCAGCCGCCCGGATCCAGCA	7
BCG <sub>A</sub> 2ab	CAACTCCACGGCGACCACCC CGATGATCTTCTGTTTGACC	8
BCG∆3ab	CACCTCGACCACGGCCAACC CCGCACCAACACCTTCCACC	9

5

Where a deletion is detected by determining the presence or absence of sequences contained within the deletion (deletion sequences), the absence of deletion sequences indicates the presence of a deletion and thus an avirulent phenotype. Conversely, the presence of deletion sequences indicates the absence of a deletion. Deletion sequences that provide suitable targets for detecting the deletions of the present invention are provided in Figures 1, 2 and 3.

15

## A) Isolation of DNA for Detection of Mycobacterium Genomic Deletions

20

In a preferred embodiment, DNA is obtained from mycobacteria. As used herein, the term "mycobacteria" refers to any bacteria of the family Mycobacteriaceae (order Actinomycetales) and includes, but is not limited to, Mycobacterium tuberculosis, Mycobacterium avium complex, Mycobacterium kansasii, Mycobacterium scrofulaceum, Mycobacterium bovis and Mycobacterium leprae. These species and groups and others are described in Baron, S., ed. Medical Microbiology, 3rd Ed. (1991) Churchill Livingstone, New York, which is incorporated herein by reference.

25

The identification of deletions using a DNA marker requires that the DNA sequence be accessible to the particular probes used or to the components of the amplification system if the DNA sequence is to be amplified. In general, this accessibility is ensured by isolating the nucleic acids from the sample.

30

A variety of techniques for extracting nucleic acids from biological samples are known in the art. For example, see those described by Sambrook et al., Molecular Cloning - A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, New

York, (1985), by Han, et al. Biochemistry, 26: 1617-1625 (1987) and by Du, et al. Bio/Technology, 10: 176-181 (1992), which are incorporated herein by reference.

Alternatively, if the sample is readily disruptable, the nucleic acid need not be purified prior to amplification by the PCR technique, *i.e.*, if the sample is comprised of cells, particularly peripheral blood lymphocytes or monocytes, lysis and dispersion of the intracellular components may be accomplished merely by suspending the cells in hypotonic buffer or boiling them in a low concentration of alkali (*i.e.* 10 mM NaOH).

In a preferred embodiment, DNA is extracted from mycobacteria as described in Example 1.

10

15

20

25

30

5

### B) Detection of Deletions Using Hybridization Probes

In one embodiment the avirulence deletions are detected by contacting DNA obtained from the mycobacterium with a probe that specifically binds an entire deletion junction region or a subsequence of that region and does not specifically bind to any other DNA sequences in the sample. Alternatively, a probe that specifically binds the entire deleted region or subsequence of that region and does not specifically bind to any other sequences in the sample is also suitable. While such probes may be proteins, oligonucleotide probes are preferred. Typically, the sequence of the oligonucleotide probe is chosen to be complementary to a select subsequence unique to the deletion junction or the deletion sequence, whose presence or absence is to be detected. Under stringent conditions the probe will hybridize with the select subsequence forming a stable duplex.

The probe is typically labeled. Detection of the label in association with the target DNA indicates either the presence or absence of the deletion. The probe may be used to detect the deletion junction or deletion sequences directly in a DNA sample without amplification of the deletion subsequences. In one embodiment, unamplified DNA sequences are probed using a Southern blot. The DNA of the sample is immobilized, on a solid substrate, typically a nitrocellulose filter or a nylon membrane. The substrate-bound DNA is then hybridized with the labeled probe under stringent conditions and non-specifically hybridized probe is washed away. Labeled probe detected in association with the immobilized mycobacterial sequences (e.g. bound to the substrate) indicates the presence of deletion sequences (e.g. BCGa1, BCGa2, or BCGa3) and therefore the absence of the deletion. Means for detecting specific DNA sequences are well known to those of skill in

10

15

20

25

30

the art. Protocols for Southern blots as well as other detection methods are provided in Maniatis, et al. Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory Press, NY (1982), which is incorporated herein by reference.

In another embodiment, the mycobacterial DNA subsequences are themselves labeled. They are then hybridized, under stringent conditions, with a probe immobilized on a solid substrate. Detection of the label in association with the immobilized probe indicates the presence or absence of the deletion.

In a preferred embodiment, the deletion junction sequences or subsequences or the deletion sequences or subsequences may be amplified by a variety of DNA amplification techniques (for example via cloning, polymerase chain reaction, ligase chain reaction, transcription amplification, etc.) prior to detection using a probe. Because the copy number of mycobacterial sequences bearing the virulence-attenuating deletions is low, the use of unamplified mycobacterial DNA results in an assay of low sensitivity. Amplification of mycobacterial DNA increases sensitivity of the assay by providing more copies of possible target subsequences. In addition, by using labeled primers in the amplification process, the mycobacterial DNA sequences are labeled as they are amplified.

# C) Selection of Probes for Detection of the Deletion Junction Sequences or the Deletion Sequences

Full length sequences are provided for the deletions BCGa1, BCGa2, and BCGa3 in Figures 1, 2 and 3 respectively. Using these sequence listings, one of skill in the art may easily determine appropriate probes or primers for the detection of the presence or absence of the deletion junctions or the deletion sequences. Generally speaking, a probe will be selected that hybridizes to the target junction sequences or deletion sequences, but not to other mycobacterial nucleic acid sequences under stringent conditions. The design of hybridization probes is well known in the art. See, for example, Sambrook et al., Molecular Cloning - A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, (1989), which is incorporated herein by reference.

In a preferred embodiment, the probe is an oligonucleotide sequence complementary to a subsequence comprising a deletion junction (e.g. BCGala, BCGalb, BCGala, BCGala, BCGala, BCGalab, BCGalab

10

15

20

25

30

sequence complementary to a subsequence of a deletion sequence (e.g. BCG $\Delta$ 1, BCG $\Delta$ 2, and BCG $\Delta$ 3). The probe preferably has destabilizing mismatches with subsequences from other regions of the mycobacterial genome.

The exact length of the probe depends on many factors including the length of conserved regions around the deletions, the degree of sequence specificity desired, and the amount of internal complementarity within the probe. Such probes are preferably 17 to 25 bases in length. One of skill will recognize that longer probes specifically hybridize at higher temperatures. Generally, stringent conditions are selected to be about 5°C to 20°C, more preferably about 10°C, lower than the thermal melting point (T<sub>m</sub>) for the specific sequence at a defined ionic strength and pH. Under stringent conditions, the probe will specifically hybridize to a nucleic acid sequence from an avirulent mycobacterium such as BCG, but not to a nucleic acid sequence from a virulent mycobacterium such as MTB or MBV. Alternatively, Under stringent conditions, the probe will specifically hybridize to a nucleic acid sequence from a avirulent mycobacterium such as MTB or MBV, but not to a nucleic acid sequence from an avirulent mycobacterium such as BCG.

Oligonucleotide probes can be prepared by any suitable method, including, for example, cloning and restriction of appropriate sequences and direct chemical synthesis by a method such as the phosphotriester method of Narang et al. Meth. Enzymol, 68: 90-99 (1979); the phosphodiester method of Brown et al., Meth. Enzymol. 68:109-151 (1979); the diethylphosphoramidite method of Beaucage et al., Tetra. Lett., 22: 1859-1862 (1981); and the solid support method of U.S. Patent No. 4,458,066.

Probe detectability may be increased by the attachment of a label. As used herein, a label is any composition detectable by spectroscopic, photochemical, biochemical, immunochemical, electrical, optical or chemical means. Useful labels in the present invention include magnetic beads (e.g. Dynabeads<sup>TM</sup>), fluorescent dyes (e.g., fluorescein isothiocyanate, texas red, rhodamine, and the like), radiolabels (e.g., <sup>3</sup>H, <sup>125</sup>I, <sup>35</sup>S, <sup>14</sup>C, or <sup>32</sup>P), enzymes (e.g., horse radish peroxidase, alkaline phosphatase and others commonly used in an ELISA), and colorimetric labels such as colloidal gold or colored glass or plastic (e.g. polystyrene, polypropylene, latex, etc.) beads.

Methods for attaching labels to probes, primers, and antibodies are well known to those of skill in the art. For example, the probe can be labeled at the 5'-end with  $^{32}P$  by incubating the probe with  $^{32}P$ -ATP and polynucleotide kinase (see Perbal, A

Practical Guide to Molecular Cloning, 2nd ed. John Wiley, N.Y. (1988)). Other labels may be joined to the probe directly or through linkers. They may be located at the ends of the probe or internally. Methods of attaching labels may be found in Connell, et al., Bio/Techniques 5: 342 (1987), U.S. Patent Nos. 4,914,210, 4,391,904 and 4,962,029, which are incorporated herein by reference. In addition, kits for labelling oligonucleotides are widely available. See, for example, Boehringer Mannheim Biochemicals (Indianapolis, IN) for "Genius" labeling kits based on dioxigenin technology and Clonetech (South San Francisco, CA) for a variety of direct and indirect oligonucleotide labeling reagents.

10

15

20

5

## D) Detection of Deletions Conferring Avirulence Through Amplification of Unique Subsequences

Deletions are particularly amenable to detection without the use of a hybridization probe. In a preferred embodiment, subsequences are amplified that include a deletion junction. The amplified deletion junction may be a "spanning" deletion junction in which case where the deletion is present (i.e. the deletion sequences are absent), the amplification product is a specific DNA incorporating the deletion junction sequence spanning the deletion (e.g. incorporating flanking sequences from both sides of the deleted sequence). Where the deletion is absent (i.e. deletion sequences are present) and primers are selected so that there are no priming sites within the deletion sequences, amplification is non-existent or alternatively provides a complex mixture of nonspecifically amplified fragments. Alternatively, amplification primers may be selected that specifically hybridize to deletion sequences, as long as they are selected to amplify sequences that are distinguishable from the sequence amplified when the deletion is present.

Alternatively, the amplification product may be subsequence of a "terminal" deletion junction in which case absence of the deletion (i.e. the deletion sequences are present) will result in the amplification of the specifically targeted nucleic acid. Conversely, where the deletions are present (i.e. the deletion sequences are absent) there will be no specific amplification of a terminal deletion junction.

30

25

Amplification products may be separated by size for characterization. Size separation may be accomplished by a variety of means known to those of skill in the art.

10

30

These methods include, but are not limited to electrophoresis, density gradient centrifugation, liquid chromatography, and capillary electrophoresis. In a preferred embodiment, the fragments are separated by agarose gel electrophoresis. The bands are then stained with a marker to visualize them such as ethidium bromide and the gel is visualized, e.g., using ultraviolet light.

As described above, an agarose gel typically shows 1 band if the deletion is present, reflecting amplification of the deletion-spanning sequence. Where the deletion is absent, amplification results in either no bands, where there are no sequences within the deletion to which the amplification primers may hybridize, or a smear where there is non-specific amplification, or a series of discrete bands distinguishable from the band representing the deletion-spanning sequence where primers are chosen that hybridize to deletion sequences.

# E) Selection of Primers for Amplification of Avirulence Deletions

15 Amplification of deletion junction sequences or subsequences or deletion sequences or subsequences may be accomplished by methods well known in the art, which include, but are not limited to polymerase chain reaction (PCR) (Innis, et al., PCR Protocols. A guide to Methods and Application. Academic Press, Inc. San Diego, (1990), which is incorporated herein by reference), ligase chain reaction (LCR) (see Wu and Wallace, Genomics, 4: 560 (1989), Landegren, et al., Science, 241: 1077 (1988) and 20 Barringer, et al., Gene, 89: 117 (1990), which are incorporated herein by reference), transcription amplification (see Kwoh, et al., Proc. Natl. Acad. Sci. (U.S.A.), 86: 1173 (1989) which is incorporated herein by reference), and self-sustained sequence replication (see Guatelli, et al., Proc. Nat. Acad. Sci. (U.S.A.), 87: 1874 (1990) which is incorporated herein by reference), each of which provides sufficient amplification so that 25 the target sequence can be detected by nucleic acid hybridization to a probe or by electrophoretic separation. Alternatively, methods that amplify the hybridization probe to detectable levels can be used, such as  $Q\beta$ -replicase amplification. See, for example, Kramer, et al. Nature, 339: 401 (1989), Lizardi, et al. Bio/Technology, 6: 1197 (1988), and Lomell, et al., Clin. Chem. 35: 1826 (1989) which are incorporated herein by reference.

In a preferred embodiment, amplification is by polymerase chain reaction using a pair of primers that flank and thereby amplify a selected deletion junction subsequence. Selection of primers is readily apparent to one of skill in the art using the sequence listings of the present invention. For example, a pair of PCR primers 5'-TCGACGATTGGCACAT-3' ( $T_m$ =55°C) and 5'-TCCCTCCCTGTATTTGTAT-3' ( $T_m$ =56°C) will amplify a 469 base pair sequence including the BCGala deletion junction, while 5'-CGTTCTTCGGAGGTTTC-3' ( $T_m$ =56°C) and 5'-GGCGGCTGGGTGGA-3' ( $T_m$ =60°C) will amplify a 471 base pair sequence including the BCGalb deletion junction.

10

15

20

25

30

5

# F) Detection of Deletions through Detection of Expression Products of Deletion Sequences

In addition to the detection of deletions by the detection of either the deletion junction sequences or the deletion sequences, one may detect the absence of the deletion by detecting the expression products of the deletion sequences. Thus, for example, where the deletion sequences express a protein, the presence of that protein indicates the absence of the deletion and thus is indicative of a virulent (non BCG-like) phenotype. Such proteins are referred to herein as "deletion polypeptides".

Means of determining proteins expressed by particular nucleic acid sequences are well known to those of skill in the art. Typically this involves determining the longest open reading frame. This may be aided by the identification of initiation sites (e.g. ribozome binding sites). The protein encoded by the largest open reading frame is determined using codon preferences for the specific organism from which the nucleic acid is obtained. The polypeptide sequence listing may then be compared against a sequence database, e.g. GenBank, to determine other sequences sharing substantial sequence identity with the calculated sequence. The expression of the protein may be verified by isolating and then sequencing proteins having the predicted length and charge characteristics.

Once deletion polypeptides are identified they may be detected by routine methods well known to those of skill in the art. Typically this involves isolating and then detecting the polypeptide. The polypeptide may be isolated by a number of means well known to those of skill in the art. This includes typical methods of protein

purification such as high performance liquid chromatography (HPLC), electrophoresis, capillary electrophoresis, hyperdiffusion chromatography, thin layer chromatography, and the like. Methods of purifying and detecting proteins are well known to those of skill in the art (see, e.g., Methods in Enzymology Vol. 182: Guide to Protein Purification, M. Deutscher, ed. Vol. 182 (1990), which is incorporated herein by reference).

Alternatively, deletion polypeptides sequences may be detected using immunoassays utilizing antibodies specific for the deletion polypeptides. The production of such antibodies and their use in immunoassays is detailed below.

#### G) Antibodies to Deletion Polypeptides

Antibodies can be raised to the polypeptides encoded by the nucleic acids corresponding to the open reading frames present in the deletion regions of the present invention (deletion polypeptides). As used herein "antibodies" include immunoglobulin or a population of immunoglobins which specifically bind to an antigen. Thus an antibody may be monoclonal or polyclonal including individual, allelic, strain, or species variants, and fragments thereof, both in their naturally occurring (full-length) forms and in recombinant forms. Additionally, antibodies can be raised to these polypeptides in either their native configurations or in non-native configurations. Anti-idiotypic antibodies may also be used.

20

25

30

15

5

10

#### 1) Antibody Production

A number of immunogens may be used to produce antibodies specifically reactive with deletion polypeptides. Recombinant polypeptides are the preferred immunogen for the production of monoclonal or polyclonal antibodies. Naturally occurring polypeptides may also be used either in pure or impure form. Synthetic peptides made using sequences described herein may also used as immunogens for the production of antibodies.

Recombinant polypeptides are expressed in eukaryotic or prokaryotic cells and purified using standard techniques. The polypeptide is injected into an animal capable of producing antibodies. Either monoclonal or polyclonal antibodies may be generated for subsequent use in immunoassays to measure the presence and quantity of the polypeptide.

2.50

5

10

15

20

25

30

Methods of producing polyclonal antibodies are known to those of skill in the art. In brief, an immunogen, preferably a purified deletion polypeptide is mixed with an adjuvant and animals are immunized with the mixture. The animal's immune response to the immunogen preparation is monitored by taking test bleeds and determining the titer of reactivity to the polypeptide of interest. When appropriately high titers of antibody to the immunogen are obtained, blood is collected from the animal and antisera are prepared. Further fractionation of the antisera to enrich for antibodies reactive to the polypeptide is performed where desired. See, e.g., Coligan (1991) Current Protocols in Immunology Wiley/Greene, NY; and Harlow and Lane (1989) Antibodies: A Laboratory Manual Cold Spring Harbor Press, NY, which are incorporated herein by reference.

Monoclonal antibodies may be obtained by various techniques familiar to those skilled in the art. Description of techniques for preparing such monoclonal antibodies may be found in, e.g., Stites et al. (eds.) Basic and Clinical Immunology (4th ed.) Lange Medical Publications, Los Altos, CA, and references cited therein; Harlow and Lane (1988) Antibodies: A Laboratory Manual CSH Press; Goding (1986) Monoclonal Antibodies: Principles and Practice (2d ed.) Academic Press, New York, NY; and particularly in Kohler and Milstein (1975) Nature 256: 495-497, which discusses one method of generating monoclonal antibodies.

Summarized briefly, this method involves injecting an animal with an immunogen. The animal is then sacrificed and cells taken from its spleen, which are then fused with myeloma cells (See, Kohler and Milstein (1976) Eur. J. Immunol. 6: 511-519, incorporated herein by reference). The result is a hybrid cell or "hybridoma" that is capable of reproducing in vitro.

Colonies arising from single immortalized cells are screened for production of antibodies of the desired specificity and affinity for the antigen, and yield of the monoclonal antibodies produced by such cells is enhanced by various techniques, including injection into the peritoneal cavity of a vertebrate host. Alternatively, one may isolate DNA sequences which encode a monoclonal antibody or a binding fragment thereof by screening a DNA library from human B cells according to the general protocol outlined by Huse et al. (1989) Science 246: 1275-1281. In this manner, the individual antibody species obtained are the products of immortalized and cloned single B

10

15

20

25

30

cells from the immune animal generated in response to a specific site recognized on the immunogenic substance.

Other suitable techniques involve selection of libraries of antibodies in phage or similar vectors. See, Huse et al. Science 246: 1275-1281 (1989); and Ward, et al. Nature 341: 544-546 (1989). The polypeptides and antibodies of the present invention are used with or without modification, including chimeric antibodies. Frequently, the polypeptides and antibodies will be labeled by joining, either covalently or non-covalently, a substance which provides for a detectable signal. A wide variety of labels and conjugation techniques are known and are reported extensively in both the scientific and patent literature. Suitable labels include radionucleotides, enzymes, substrates, cofactors, inhibitors, fluorescent moieties, chemiluminescent moieties, magnetic particles, and the like. Patents, teaching the use of such labels include U.S. Patent Nos. 3,817,837; 3,850,752; 3,939,350; 3,996,345; 4,277,437; 4,275,149; and 4,366,241. Also, recombinant immunoglobulins may be produced. See, Cabilly, U.S. Patent No. 4,816,567; and Queen et al. Proc. Nat'l Acad. Sci. USA 86: 10029-10033 (1989).

Antibodies, including binding fragments and single chain versions, against predetermined fragments of deletion polypeptides can be raised by immunization of animals with conjugates of the fragments with carrier proteins as described above. Monoclonal antibodies are prepared from cells secreting the desired antibody. These antibodies can be screened for binding to normal or defective polypeptides, or screened for agonistic or antagonistic activity, e.g., mediated through a receptor. These monoclonal antibodies will usually bind with at least a  $K_D$  of about 1 mM, more usually at least about 300  $\mu$ M, and most preferably at least about 0.1  $\mu$ M or better.

The antibodies of this invention can also be used for affinity chromatography in isolating deletion polypeptides. Columns can be prepared where the antibodies are linked to a solid support, e.g., particles, such as agarose, Sephadex, or the like, where a bacterial lysate, or recombinant cell lysate is passed through the column, washed, and treated with increasing concentrations of a mild denaturant, whereby purified deletion polypeptides are released.

10

15

20

25

30

The antibodies can be used to screen expression libraries for particular expression products. Usually the antibodies in such a procedure will be labeled with a moiety allowing easy detection of presence of antigen by antibody binding.

In a preferred embodiment, antibodies to deletion polypeptides are used for the identification of cell populations expressing the polypeptides. By assaying the expression products of cells expressing the polypeptides it is possible to diagnose bacterial infections.

Antibodies raised against each polypeptide are useful to raise anti-idiotypic antibodies. These will be useful in detecting or diagnosing various immunological conditions related to the presence of the respective antigens.

#### 2) Immunoassays

A particular deletion polypeptide can be measured by a variety of immunoassay methods. For a review of immunological and immunoassay procedures in general, see Stites and Terr (eds.) 1991 Basic and Clinical Immunology (7th ed.). Moreover, the immunoassays of the present invention can be performed in any of several configurations, e.g., those reviewed in Maggio (ed.) (1980) Enzyme Immunoassay CRC Press, Boca Raton, Florida; Tijan (1985) "Practice and Theory of Enzyme Immunoassays," Laboratory Techniques in Biochemistry and Molecular Biology, Elsevier Science Publishers B.V., Amsterdam; and Harlow and Lane Antibodies, A Laboratory Manual, supra, each of which is incorporated herein by reference. See also Chan (ed.) (1987) Immunoassay: A Practical Guide Academic Press, Orlando, FL; Price and Newman (eds.) (1991) Principles and Practice of Immunoassays Stockton Press, NY; and Ngo (ed.) (1988) Non-isotopic Immunoassays Plenum Press, NY.

Immunoassays for measurement of deletion polypeptides can be performed by a variety of methods known to those skilled in the art. In brief, immunoassays to measure the protein can be, e.g., competitive or noncompetitive binding assays. In competitive binding assays, the sample to be analyzed competes with a labeled analyte for specific binding sites on a capture agent bound to a solid surface. Preferably the capture agent is an antibody specifically reactive with a deletion polypeptide produced as described above. The concentration of labeled analyte bound to the capture agent is inversely proportional to the amount of free analyte present in the sample.

WO 96/25519 PCT/US96/01938

26

In a competitive binding immunoassay, the deletion polypeptide present in the sample competes with labelled protein for binding to a specific binding agent, for example, an antibody specifically reactive with a particular deletion polypeptide. The binding agent is, e.g., bound to a solid surface to produce separation of bound labelled polypeptide from the unbound labelled polypeptide. Alternately, the competitive binding assay may be conducted in liquid phase and any of a variety of techniques known in the art may be used to separate the bound labelled protein from the unbound labelled protein. Following separation, the amount of bound labeled protein is determined. The amount of polypeptide present in the sample is inversely proportional to the amount of labelled polypeptide binding.

5

10

15

20

25

30

Alternatively, a homogenous immunoassay may be performed in which a separation step is not needed. In these immunoassays, the label on the protein is altered by the binding of the protein to its specific binding agent. This alteration in the labelled protein results in a decrease or increase in the signal emitted by label, so that measurement of the label at the end of the immunoassay allows for detection or quantitation of the polypeptide.

Deletion polypeptides may also be detected by a variety of noncompetitive immunoassay methods. For example, a two-site, solid phase sandwich immunoassay may be used. In this type of assay, a binding agent for the protein, for example an antibody, is attached to a solid support. A second protein binding agent, which is also an antibody, and which binds the protein at a different site, is labelled. After binding at both sites on the protein, the unbound labelled binding agent is removed and the labelled binding agent bound to the solid phase is measured. The amount of labelled binding agent bound is directly proportional to the amount of polypeptide in the sample.

Western blot analysis can be used to determine the presence of a deletion polypeptide in a sample. Electrophoresis is carried out, for example, on a bacterial sample suspected of containing the deletion polypeptide. Following electrophoresis to separate the proteins, and transfer of the proteins to a suitable solid support such as a nitrocellulose filter, the solid support is incubated with an antibody reactive with the protein. This antibody is labelled, or alternatively may be it is detected by subsequent incubation with a second labelled antibody that binds the primary antibody.

10

15

20

25

30

The immunoassay formats described above employ labelled assay components. The label can be in a variety of forms as described above. The choice of label depends on sensitivity required, ease of conjugation with the compound, stability requirements, and available instrumentation. For a review of various labelling or signal producing systems which may be used, see U.S. Patent No. 4,391,904, which is incorporated herein by reference.

Antibodies reactive with a particular protein can also be measured by a variety of immunoassay methods. For a review of immunological and immunoassay procedures applicable to the measurement of antibodies by immunoassay techniques, see Stites and Terr (eds.) Basic and Clinical Immunology (7th ed.) supra; Maggio (ed.) Enzyme Immunoassay, supra; and Harlow and Lane Antibodies, A Laboratory Manual, supra.

In brief, immunoassays to measure antisera reactive with polypeptides include competitive and noncompetitive binding assays. In competitive binding assays, the sample analyte competes with a labeled analyte for specific binding sites on a capture agent bound to a solid surface. Preferably the capture agent is a purified recombinant deletion polypeptide as described above. Other sources of polypeptides, including isolated or partially purified naturally occurring protein, can also be used. Noncompetitive assays are typically sandwich assays, in which the sample analyte is bound between two analyte-specific binding reagents. One of the binding agents is used as a capture agent and is bound to a solid surface. The second binding agent is labelled and is used to measure or detect the resultant complex by visual or instrument means. A number of combinations of capture agent and labelled binding agent can be used. A variety of different immunoassay formats, separation techniques and labels can be also be used similar to those described above for the measurement of deletion polypeptides.

## II. Preparation of Deletion-Containing Mycobacteria

Mycobacteria containing specific deletions may be prepared by using methods of homologous recombination well known to those of skill in the art. In brief, homologous recombination is a natural cellular process which results in the scission of two nucleic acid molecules having identical or substantially similar (i.e. "homologous") sequences, and the ligation of the two molecules such that one region of each initially

10

15

20

25

30

present molecule is now ligated to a region of the other initially present molecule (Sedivy, *Bio/Technol.*, 6: 1192-1196 (1988).

Homologous recombination is exploited by a number of various methods of "gene targeting" well known to those of skill in the art. (see, for example, Mansour et al. Nature, 336: 348-352 (1988); Capecchi Trends Genet. 5: 70-76 (1989); Capecchi Science 244: 1288-1292 (1989); Capecchi et al. pages 45-52 In: Current Communications in Molecular Biology, Capecchi, M.R. (ed.), Cold Spring Harbor Press, Cold Spring Harbor, N.Y. (1989); Frohman et al. Cell 56: 145-147 (1989)). Some approaches focus on increasing the frequency of recombination between two DNA molecules by treating the introduced DNA with agents which stimulate recombination (e.g. trimethylpsoralen, UV light, etc.), however, most approaches utilize various combinations of selectable markers to facilitate isolation of the transformed cells.

One such selection method is termed positive/negative selection (PNS) (Thomas and Cappechi Cell 51: 503-512 (1987)). This method involves the use of two selectable markers: one a positive selection marker such as the bacterial gene for neomycin resistance (neo'); the other a negative selection marker such as the herpes virus thymidine kinase (tk) gene. Neo' confers resistance to the drug G-418, while herpes tk renders cells sensitive to the nucleoside analog gangcyclovir (GANC) or 1-(2-deoxy-2-fluoro-b-d-arabinofuranosyl)-5-iodouracil (FIAU). The DNA encoding the positive selection marker in the transgene (e.g. neo<sup>R</sup>) is generally linked to an expression regulation sequence that allows for its independent transcription in mycobacteria. It is flanked by first and second sequence portions of at least a part of the deletion or deletion flanking sequences.

These first and second sequence portions target the transgene to a specific nucleotide sequence. A second independent expression unit capable of producing the expression product for a negative selection marker, e.g. for herpes virus tk is positioned adjacent to or in close proximity to the distal end of the first or second portions of the first DNA sequence. Upon transfection, some of the mycobacteria incorporate the transgene by random integration, others by homologous recombination between the endogenous allele and sequences in the transgene. As a result, one copy of the targeted nucleic acid is disrupted by homologous recombination with the-transgene with simultaneous loss of the sequence encoding herpes tk gene. Random integrants, which

10

15

20

25

30

occur via the ends of the transgene, contain herpes tk and remain sensitive to GANC or FIAU. Therefore, selection, either sequentially or simultaneously with G418 and GANC enriches for transfected mycobacteria containing the transgene integrated into the genome by homologous recombination.

Methods of homologous recombination in mycobacteria are described in greater detail by Ganjam et al. Proc. Natl. Acad. Sci. USA, 88: 5433-5437 (1991) and Aldovini et al., J. Bacteriol., 175: 7282-7289 (1993) which are incorporated herein by reference.

## III. Screening for Drug Susceptibility/Therapeutics

The expression products of the open reading frames in the BCGa1, BCGa2, and BCGa3 deletions of the present invention are targets for anti-mycobacterial drugs. To determine particularly suitable drug targets, open reading frames and surrounding expression control sequences are introduced into avirulent strains of mycobacteria, alone or in combination with other open reading frame regions to determine which regions are critical for virulence. Once particular genes are identified as critical for virulence, anti-mycobacterial agents are designed to inhibit expression of the critical genes, or to attack the critical gene products. For instance, antibodies are generated against the critical gene products and used as prophylactic or therapeutic agents. Alternatively, small molecules can be screened for the ability to selectively inhibit expression of the critical gene products, e.g., using recombinant expression systems which include the gene's endogenous promoter. These small molecules are then used as therapeutics, or prophylactic agents to inhibit mycobacterial virulence.

In another embodiment, anti-mycobacterial agents which render a virulent mycobacterium avirulent can be operably linked to expression control sequences and used to transform a virulent mycobacterium. Such anti-mycobacterial agents inhibit the replication of a specified mycobacterium upon transcription or translation of the agent in the mycobacterium.

Such transformed mycobacteria are useful as vaccine components, and as components of immunological infectivity assays. For instance, an animal's blood can be monitored for the presence of anti-mycobacterial antibodies using the procedures described herein, using transformed avirulent mycobacterial components in various

WO 96/25519

5

10

15

20

25

immunological assays. Anti-mycobacterial agents useful in this invention include, without limitation, antisense genes, ribozymes, decoy genes, transdominant proteins and suicide genes.

An antisense nucleic acid is a nucleic acid that, upon expression, hybridizes to a particular mRNA molecule, to a transcriptional promoter or to the sense strand of a gene. By hybridizing, the antisense nucleic acid interferes with the transcription of a complementary DNA, the translation of an mRNA, or the function of a catalytic RNA. Antisense molecules useful in this invention include those that hybridize to gene transcripts in the region of the deletions of the invention, particularly deletion region 1.

A ribozyme is a catalytic RNA molecule that cleaves other RNA molecules having particular nucleic acid sequences. Ribozymes useful in this invention are those that cleave deletion gene transcripts. Examples include hairpin and hammerhead ribozymes.

A decoy nucleic acid is a nucleic acid having a sequence recognized by a regulatory DNA binding protein (i.e., a transcription factor). Upon expression, the transcription factor binds to the decoy nucleic acid, rather than to its natural target in the genome. Useful decoy nucleic acid sequences include any sequence to which a transcription factor binds in the deletion regions of the present invention.

A transdominant protein is a protein whose phenotype, when supplied by transcomplementation, will overcome the effect of the native form of the protein. For instance, an avirulent mycobacterium can be rendered virulent by introducing transdominant proteins from deletion region 1.

A suicide gene produces a product which is cytotoxic. In the vectors of the present invention, a suicide gene is operably linked to an inducible expression control sequences which is stimulated upon infection of a cell by a mycobacterium.

## IV. Use of Expressed "Deletion Proteins" in a Vaccine

The deletion polypeptides encoded by the open reading frames in BCGa1, BCGa2, and BCGa3 may be recombinantly expressed and used as components of immunological assays as described above or in vaccines. Expression of polypeptides

10

15

20

25

30

encoded by the open reading frames of the BCGa1, BCGa2, or BCGa3 deletions may be accomplished by means well known to those of skill in the art.

In brief, the expression of natural or synthetic nucleic acids encoding deletion polypeptides will typically be achieved by operably linking the DNA or cDNA to a promoter (which is either constitutive or inducible), followed by incorporation into an expression vector. The vectors can be suitable for replication and integration in either prokaryotes or eukaryotes. Typical expression vectors contain transcription and translation terminators, initiation sequences, and promoters useful for regulation of the expression of polynucleotide sequence encoding deletion polypeptides.

To obtain high level expression of a cloned gene, such as those polynucleotide sequences encoding deletion polypeptides, it is desirable to construct expression plasmids which contain, at the minimum, a promoter to direct transcription, a ribosome binding site for translational initiation, and a transcription/translation terminator. The expression vectors may also comprise generic expression cassettes containing at least one independent terminator sequence, sequences permitting replication of the plasmid in both eukaryotes and prokaryotes, i.e., shuttle vectors, and selection markers for both prokaryotic and eukaryotic systems. For detailed techniques employed in the recombinant expression of deletion proteins see, for example, Sambrook, et al., Molecular Cloning: A Laboratory Manual (2nd Ed., Vols. 1-3, Cold Spring Harbor Laboratory (1989)), Methods in Enzymology, Vol. 152: Guide to Molecular Cloning Techniques (Berger and Kimmel (eds.), San Diego: Academic Press, Inc. (1987)), or Current Protocols in Molecular Biology, (Ausubel, et al. (eds.), Greene Publishing and Wiley-Interscience, New York (1987), all of which are incorporated herein by reference.

The expressed deletion polypeptides may be used in a variety of assays. For example, the deletion polypeptides can be used as reagents in immunoblot assays to test whether a patient was previously exposed to virulent mycobacteria (i.e., to test whether the patient has antibodies to the deletion polypeptide). These assays have the advantage of discriminating between previous exposure to an avirulent mycobacterium (e.g., one used in a vaccine) and exposure to a virulent mycobacterium. Thus, vaccinated individuals can be tested for antibodies to the virulent mycobacterium without regard to whether the patient has been vaccinated with an avirulent mycobacterium.

10

15

20

25

30

The deletion polypeptides can also be used as antigenic vaccine components to direct antibodies to elements which are critical for virulence. These polypeptides can be added to existing vaccines (e.g., those based upon avirulent mycobacteria and which lack the deletion polypeptide) to supplement the range of antigenicity conferred by the vaccine, or they may be used apart from other mycobacterial antigens. The vaccines of the invention contain as an active ingredient an immunogenically effective amount of a deletion polypeptide or of a recombinant vector which includes the deletion polypeptide. The immune response can include the generation of antibodies; activation of cytotoxic T lymphocytes (CTL) against cells presenting peptides derived from the polypeptides or other mechanisms well known in the art. See e.g. Paul Fundamental Immunology Third Edition published by Raven press New York (incorporated herein by reference) for a description of immune response. Useful carriers are well known in the art, and include, for example, thyroglobulin, albumins such as human serum albumin, tetanus toxoid, and polyamino acids such as poly(D-lysine:D-glutamic acid). The vaccines can also contain a physiologically tolerable (acceptable) diluent such as water, phosphate buffered saline, and further typically include an adjuvant. Adjuvants such as incomplete Freund's adjuvant, aluminum phosphate, aluminum hydroxide, or alum are materials well known in the art.

The compositions are suitable for single administrations or a series of administrations. When given as a series, inoculations subsequent to the initial administration are given to boost the immune response and are typically referred to as booster inoculations.

The vaccine compositions of the invention are intended for parenteral, topical, oral or local administration. Preferably, the pharmaceutical compositions are administered parenterally, e.g., intravenously, subcutaneously, intradermally, or intramuscularly. Thus, the invention provides compositions for parenteral administration that comprise a solution of the agents described above dissolved or suspended in an acceptable carrier, preferably an aqueous carrier. A variety of aqueous carriers may be used, e.g., water, buffered water, 0.4% saline, 0.3% glycine, hyaluronic acid and the like. These compositions may be sterilized by conventional, well known sterilization techniques, or may be sterile filtered. The resulting aqueous solutions may be packaged for use as is, or lyophilized, the lyophilized preparation being combined with a sterile

- Z

5

10

15

20

25

30

solution prior to administration. The compositions may contain pharmaceutically acceptable auxiliary substances as required to approximate physiological conditions, such as pH adjusting and buffering agents, tonicity adjusting agents, wetting agents and the like, for example, sodium acetate, sodium lactate, sodium chloride, potassium chloride, calcium chloride, sorbitan monolaurate, triethanolamine oleate, etc.

For solid compositions, conventional nontoxic solid carriers may be used which include, for example, pharmaceutical grades of mannitol, lactose, starch, magnesium stearate, sodium saccharin, talcum, cellulose, glucose, sucrose, magnesium carbonate, and the like. For oral administration, a pharmaceutically acceptable nontoxic composition is formed by incorporating any of the normally employed excipients, such as those carriers previously listed, and generally 10-95% of active ingredient and more preferably at a concentration of 25%-75%.

For aerosol administration, the polypeptides are preferably supplied in finely divided form along with a surfactant and propellant. The surfactant should be nontoxic, and preferably soluble in the propellant. Representative of such agents are the esters or partial esters of fatty acids containing from 6 to 22 carbon atoms, such as caproic, octanoic, lauric, palmitic, stearic, linoleic, linolenic, olesteric and oleic acids with an aliphatic polyhydric alcohol or its cyclic anhydride. Mixed esters, such as mixed or natural glycerides may be employed. A carrier can also be included, as desired, as with, e.g., lecithin for intranasal delivery.

The amount of vaccine administered to the patient will vary depending upon the composition being administered, the physiological state of the patient and the manner of administration.

Live attenuated recombinant viruses which include the deletion polypeptide, such as recombinant vaccinia or adenovirus vectors, are convenient alternatives as vaccines because they are inexpensive to produce and are easily transported and administered. Vaccinia vectors and methods useful in immunization protocols are described, for example, in U.S. Patent No. 4,722,848, incorporated herein by reference.

Deletion sequences and subsequences of this invention may also be used in methods of genetic immunization. Briefly, genetic immunization involves transfecting

10

15

20

25

30

cells in vivo with nucleic acids encoding pathogen specific antigens. The transformed host cells then express the antigen thereby stimulating the host immune system.

In the present invention, antigen-encoding deletion region sequences are used to transform mammalian host cells thereby resulting in the expression of the antigen by the host. This provokes an immune response by the host against the expressed antigen thereby conferring immunity on the host. Methods of genetic immunization are well known to those of skill in the art (see, e.g., Wang et al. Proc. Natl. Acad. Sci. USA, 90: 4156-4160 (1993); Ulmer et al., Science, 259: 1745-1749 (1993); Fynan et al. DNA Cell Biol., 12: 785-789 (1993); Fynan et al. Proc. Natl. Acad. Sci. USA, 90: 11478-11482 (1993); Robinson et al. Vaccine, 11: 957-960 (1993); and Martinon et al. Eur. J. Immunol., 23: 1719-1722 (1993), which are incorporated herein by reference.

# VI. Use of Promoters within Deletion Sequences for Expression of Recombinant Proteins

Bacille Calmette-Guérin (BCG) contains all three deletions (BCGa1, BCGa2, and BCGa3) and yet is able to grow and reproduce indicating that the sequences contained within the deletion are not essential for bacterial viability. These deletion regions therefore make good target sites for the insertion of heterologous DNA as mycobacteria are tolerant of disruption of the native genome in these regions. The BCGa1, BCGa2, and BCGa3 deletion regions therefore provide suitable target sites for the incorporation of expression cassettes and the subsequent expression of exogenous gene products. The expression cassettes typically comprise a nucleic acid sequence under the control of a promoter. The promoter may be either constitutive or inducible. The cassette may additionally comprise a selectable marker such as an antibiotic resistance gene, a gene encoding a fluorescent marker (e.g. green fluorescent protein), or a gene encoding an enzymatic marker (e.g. B-galactosidase).

Alternatively, genes under the control of endogenous promoters may be used as well. In one embodiment, reporter genes under the control of endogenous promoters found within the deletion sequences may be inserted at the deletion sites. These reporter genes may be utilized as an assay for antimycobacterial compounds that act by inhibiting transcription or translation of deletion sequences. Assaying for the

\*\*\*

nz

5

15

20

25

30

reporter gene product in the presence of an antimycobacterial compound provides a measure of efficacy of that compound in upregulating or downregulating deletion sequence genes. Methods of use of mycobacterial reporter gene assays to screen for drug activity are described by Cooksey et al., Antimicrob. Agents Chemother., 37: 1348-1352 (1993), and Jacobs et al., Science, 260: 819-822 (1993) which are incorporated herein by reference.

### **EXAMPLES**

The following examples are offered by way of illustration, not by way of limitation.

### Example 1

## Identification of Virulence-Attenuating Deletions

### **Bacterial Culture**

All strains of Mycobacteria used in this study were maintained in 7H9 (Difco, Detroit Michigan, USA) media supplemented with OADC (BBL) or were grown on 7H11 agar supplemented with oleic acid albumin dextrose complex (OADC). Escherichia coli (strain DH5 $\alpha$  or NM554) was used as a host for all recombinant plasmids and cosmids. E. coli was maintained in LB medium with or without agar. Carbenicillin (100  $\mu$ g/ml) was used in place of ampicillin for the selection of all E. coli plasmids.

### Extraction of High Molecular Weight DNA

High molecular weight chromosomal DNA was prepared by diluting a late log phase culture of the respective mycobacterium 1:10 into a liter of 7H9 medium containing 1.5% glycine and continuing growth for 4 to 5 days. The cells were then harvested by centrifugation, washed once in TE (pH 8.0) and resuspended in 4 ml of 25% sucrose in 10X TE. 100  $\mu$ g of lysozyme was added and the preparation was incubated at 37°C for 2 hr followed by the addition of 100  $\mu$ g of proteinase K and sarkosyl to a concentration of 1% weight/volume. Following overnight incubation at 65°C the mixture was extracted 4 times with chloroform isoamyl alcohol 24:1, once with phenol/chloroform (1:1), and twice again with chloroform isoamyl alcohol. The resulting high molecular weight DNA was then run on a CsCl gradient as described by

Hull et al. Infect. Immun., 33: 933-938 (1981), which is incorporated herein by reference, and subsequently dialyzed against 4 changes of TE. BCG DNA was physically sheared by passage through a 22 gauge needle until an average size of 3-10 kb was obtained (20-25 passages). This DNA was then biotinylated using photobiotin (Clonetech, Palo Alto, California, USA) according to the method of Straus and Ausubel, Proc. Natl. Acad. Sci. USA, 87: 1889-1893 (1990), which is incorporated herein by reference.

### **DNA Subtraction**

5

10

15

20

25

30

DNA subtraction was carried out between virulent *M. tuberculosis* H37Rv and avirulent BCG. H37R chromosomal DNA was selected because it was the most readily available chromosomal DNA from a virulent strain. In addition, *M. bovis* and *M tuberculosis* H37Rv are highly homologous.

M. bovis/M. tuberculosis specific probes were generated by the method of Straus and Ausubel, supra. with the following modifications. Sheared and biotinylated BCG DNA was used in a 10:1 excess for each round of subtraction. Wild type M. tuberculosis H37Rv DNA was digested with Sau3A to an average size of 1 kb. Hybridization conditions were 1M NaCl and 65 °C for 18 hours. Following five cycles (successive denaturation and reassociations) of subtraction, Sau3A1 adaptors (GACACTCTCGAGACATCACCGTCC and GATCGGACGGTGATGTCTCGAGAGTG were ligated to the subtraction product and amplified in a PCR reaction for 35 cycles (30 sec at 95°C, 30 sec at 55°C, and 3 min at 72°C). The M. tuberculosis/M. bovis specific probes were radiolabeled by using one strand of the adaptor (GACACTCTCGAGACATCACCGTCC) as a primer and labeling with <sup>32</sup>P dCTP using the Klenow fragment of DNA polymerase.

An M. bovis cosmid library was constructed in the BamH1 site of sCOS (Stratagene, La Jolla California, USA) with subsequent in vitro packaging and infection of E. coli strain NM554 (Stratagene). 600 colonies were picked to Nytran circular membranes and the membranes prepared according to the method of Grunstein and Hogness, Proc. Natl. Acad. Sci. USA, 72: 3961 (1975), which is incorporated herein by reference. These filters were then probed using the BCG subtracted probe and positive clones selected for further analysis. Cosmid DNA was prepared from selected clones by the method of Birnboim and Doly, Nucleic Acids. Res., 7: 1513 (1973) which is

10

15

20

25

30

incorporated herein by reference. Restriction fragments that hybridize with the MTB/MBV specific probe were further subcloned into pGEM7z or pGEM5z (Promega, Madison, Wisconsin, USA) for deletion analysis.

Plasmid DNA for DNA sequencing was prepared using Qiagen minicolumns (Qiagen Inc. Chatsworth California, USA) and sequenced by the method of Henikoff, Gene, 28: 351-359 (1984), which is incorporated herein by reference, using the Erase A Base System (Promega). DNA sequencing reactions were run using a Perkin Elmer 9600 thermocycler and analyzed on an automated ABI sequencer. Analysis and assembly of contiguous DNA sequence was done using the ABI analysis software and SeQuencher sequence analysis software by Gene Clones Corp (Ann Arbor, Michigan, USA).

### Deletion Region 1 (BCGA1)

Sequence analysis of over 16 kb of MBV region 1 and homologous regions in BCG revealed the precise junctions for the deletion in BCG. Eight open reading frames were identified with codon usage biases matching that of known MTB and MBV genes (see map Figure 4). The potential start and stop codons and predicted maximum protein coding capacity are listed in Figure 4. Consensus ribosomal binding site sequences were found near potential start codons for seven of eight open reading frames. TBLASTN and FASTA sequence homology analysis with each potential ORF-encoded protein revealed significant homologies for 3 of 8 open reading frames in region 1.

Most notable is the ORF1C homology to an unpublished and uncharacterized sequence listed in Genbank as M. tuberculosis antigen esat6. A 65 base pair repeated overlapping (repeated ~2 1/2 times) sequence was also recognized within the ORF1C (esat6) open reading frame. Also noteworthy are the significant homologies identified between ORF1H and bacterial serine proteases including B. subtilus subtilisin. Of the eight recognized open reading frames, four (ORFs 1B, 1C, 1D, and 1E) are located entirely within the 9 kb region deleted in BCG. One ORF traverses the BCG deletion junction in virulent M. bovis.

DNA probes from the 9 kb deletion in region 1 demonstrated that this region is absent in all BCG substrains and present in all virulent MBV and MTB strains tested. Furthermore, restriction fragment patterns observed in Southern blot analysis

10

15

20

25

30

with region 1 probes are non-polymorphic and identical in virulent MBV and MTB. This region has far fewer direct and indirect repeats than the regions 2 (BCG $\Delta$ 2) and 3 (BCG $\Delta$ 3) characterized below.

The sequence of a small region, estimated to be less than 20 bp between basepair coordinates 10654 and 10664 in region 1 has been recalcitrant to automated sequencing. Therefore, pending sequence confirmation, the base pair coordinates given in the region 1 map (Figure 4) are approximations. The precise sequence determination is likely to effect the Orf1E open reading frame.

### Deletion Region 2 (BCGA2)

Sequence analysis of over 15 kb of MBV region 2 and homologous regions in BCG revealed the precise junctions for an 11 kb deletion in BCG. Thirteen open reading frames were identified with codon usage biases matching that of known MTB and MBV genes (see map Figure 5). The potential start and stop codons and predicted maximum protein coding capacity are also shown in Figure 5. Candidate consensus sequences resembling ribosomal binding sites were found near potential start codons for eight open reading frames. Of the thirteen open reading frames recognized in BCGA2, nine are located entirely within the 11 kb region deleted in most BCG strains while ORF2B2 and ORF2I traverse the deletion junctions.

TBLASTN and FASTA sequence homology analysis with each potential ORF-encoded protein revealed significant homologies for five open reading frames in BCGa2. A protein encoded by ORF2C exhibits striking similarity to the E. coli iciA protein which is thought to play a role in inhibiting and regulating the initiation of chromosomal replication. The iciA protein product is a member of the large LysR family of transcriptional regulatory proteins. Orf2F is highly homologous to an S. typhimurium ribonucleotide diphosphate reductase and a region of the E. coli and S. typhimurium proUVWX operon. Orf2H was found to have significant homology to E. coli and S. typhimurium permeases involved in aromatic amino acid transport and a eukaryotic cell retroviral receptor.

The Orf2G encoded protein was identical to the MTB mpt64 gene previously thought to encode a secreted antigen which is specifically expressed by MTB

10

15

20

25

30

and not BCG strains. Recent analysis of mpt64 expression revealed that three BCG substrains do express mpt64 (Moreau, Tokyo, Russian). Probes specific for mpt64 or other non-repetitive parts of region 2 hybridized to all MTB strains tested and the same three BCG substrains shown to express mpt64. Of interest is the finding that these three BCG substrains are derived from the original Pasteur strain prior to 1925. The current Pasteur strain and all strains derived from the original Pasteur strain after 1925, including the Connaught strain used in the subtractive analysis in this study, are deleted in the 11 kb DNA segment contained within BCGA2. These data indicate that an additional mutational event deleting the 11 kb segment of region 2, occurred in the BCG Pasteur strain sometime after 1925.

Southern blot analysis with probes from different segments of region 2 revealed a repetitive element located within a 2 kb segment (8-10 kb) of region 2. This repetitive element is ubiquitous in all tubercle bacilli tested. This element provides a marker suitable for RFLP analysis of mycobacterial strains.

### Deletion Region 3 (BCGA3)

Sequence analysis of the almost 11 kb region 3 sequence and comparison to a homologous region in BCG precisely identified the deletion junctions for BCG. Twelve potential open reading frames were recognized in the region 3 sequence, seven of which are entirely located within the 9 kb region deleted in BCG. At least 9 ORFs in BCGA3 exhibit codon usage preferences comparable to that of the tubercle bacilli. Sequence homology analysis of presumptive protein sequences encoded by six open reading frames in region 3 revealed highly significant homology to listed sequences. Orfs3B, 3D, and 3E exhibit homology to phage sequences, suggesting a phage derivation for 4 or more kb of DNA in region 3. Homology to putative open reading frames in two M. leprae cosmids was also observed including homology to a putative bid gene encoding a protein involved in biotin synthesis. Also of interest was homology between ORF3A and an MTB sequence (mce) associated with cell invasion and intracellular survival.

Southern blot analysis with segments of region 3 deleted in BCG revealed that prototype lab strains of virulent MBV and MTB all carry deletion region 3 DNA. However, clinical isolates from PHRI are highly polymorphic or deleted in region 3.

This region contains many large direct and indirect repeats and, as mentioned above, at least 2 ORFs are homologous to phage sequences including homology to DNA invertases or recombinases. The repetitive nature of this region and the possible presence of a DNA recombinase could explain the polymorphisms observed in this region.

5

The sequence of a small region, estimated to be much less than 200 bp and located close to 9400 bp in Figure 3, was recalcitrant to automated sequencing and remains to be determined. Therefore, the base pair coordinates given in the region 3 map (Figure 6) 3' to the 9kb marker are approximations. The precise sequence determination of region is likely to effect the length of open reading frames 3H and 3L.

10

15

The foregoing subtractive analysis identified 3 regions in virulent M. bovis and M. nuberculosis prototype strains which are deleted in the avirulent BCG strain. The deletion located in region 2 may not have arisen in the original BCG Pasteur strain as this region is only deleted in strains derived from the original Pasteur strain after 1925. Region 3 is present in virulent MTB and MBV lab prototype strains (H37Rv, Erdman) and is highly polymorphic and at least partially deleted in the majority of MTB clinical isolates tested. Region 1 is apparently conserved and intact in all virulent MBV and MTB strains tested to date while all avirulent BCG strains tested to date are missing approximately 9kb from region 1.

### 20

# Screening and Identification of an Avirulent Mycobacterium

Example 2

The <sup>32</sup> P labeled subtraction probe obtained in Example 1, was used to probe EcoRI and BamHI restricted chromosomal DNAs from BCG Connaught, *Mycobacterium bovis*, and various strains of *Mycobacterium tuberculosis* in a Southern blot. The hybridization was performed at 70°C in 6X SSC overnight.

25

The resulting Southern blot is illustrated in Figure 8. The probe showed no labeling of BCG reflecting the presence of all three deletions, while the other strains were labeled.

30

The above examples are provided to illustrate the invention but not to limit its scope. Other variants of the invention will be readily apparent to one of ordinary skill in the art and are encompassed by the appended claims. All publications, patents, and patent applications cited herein are hereby incorporated by reference.

### WHAT IS CLAIMED IS:

- 1. A marker for an avirulent mycobacterium, said marker comprising a first nucleic acid that specifically hybridizes under stringent conditions with a second nucleic acid or a complement of said second nucleic acid where said second nucleic acid or complement of said second nucleic acid is selected from the group consisting of BCGala, BCGalb, BCGala, BCGala, BCGala, BCGalab, BCG
- 2. The marker of claim 1, wherein said marker specifically hybridizes under stringent conditions to a nucleic acid from BCG, but not to a nucleic acid from Mycobacterium tuberculosis or Mycobacterium bovis, or where said marker specifically hybridizes under stringent conditions to a nucleic acid from Mycobacterium tuberculosis or Mycobacterium bovis, but not to a nucleic acid from BCG.
- 3. The marker of claim 2, wherein said marker comprises a subsequence of a nucleic acid where said nucleic acid is selected from the group consisting of BCGa1a, BCGa1b, BCGa2a, BCGa2b, BCGa3a, BCGa3b, BCGa1ab, BCGa2ab, BCGa3ab, BCGa1, BCGa2, and BCGa3.
- 4. The marker of claim 2, wherein said marker is selected from the
   group consisting of BCGΔ1a, BCGΔ1b, BCGΔ2a, BCGΔ2b, BCGΔ3a, BCGΔ3b,
   BCGΔ1ab, BCGΔ2ab, BCGΔ3ab, BCGΔ1, BCGΔ2, and BCGΔ3.
  - 5. The marker of claim 2, wherein said marker comprises a nucleic acid having at least 90 percent sequence identity with a sequence selected from the group consisting of BCGala, BCGalb, BCGala, BCGala, BCGala, BCGalab, BCGalab,
- 1 6. The marker of claim 2, wherein said marker comprises a radioactive nucleotide probe.

1 The marker of claim 2, wherein said subsequence is a sequence 7. selected from an open reading frame of a deletion, said deletion being selected from the 2 group consisting of BCGa1, BCGa2, BCGa3. 3 1 A polypeptide encoded by a subsequence of a deletion sequence 8. selected from the group consisting of BCGa1, BCGa2, and BCGa3. 2 1 The polypeptide of claim 8, wherein the subsequence is selected 9. from an open reading frame (ORF) of a deletion, said deletion being selected from the 2 group consisting of BCGa1, BCGa2, BCGa3. 3 1 An antibody that binds specifically to the polypeptide of claim 8. 10. 1 A recombinant cell comprising a first nucleic acid that hybridizes 11. under stringent conditions with a second nucleic acid or a complement of said second 2 nucleic acid where said second nucleic acid or complement of said second nucleic acid is 3 selected from the group consisting of BCGala, BCGalb, BCGala, BCGalb, BCGala, 4 BCGa3b, BCGa1ab, BCGa2ab, BCGa3ab, BCGa1, BCGa2, and BCGa3. 5 1 The recombinant cell of claim 11, wherein the cell is a 12. 2 Mycobacterium. 1 The cell of claim 11, wherein the cell expresses a polypeptide 13. encoded by an intact open reading frame from BCGa1, BCGa2, and BCGa3. 2 1 The cell of claim 11, wherein said cell is a mycobacterium having 14. one or more deletions in the genomic regions selected from the group consisting of 2 BCGa1, BCGa2, and BCGa3, wherein said deletions result in the attenuation of an 3 otherwise virulent strain of mycobacterium and wherein said deletions are present in up 4 5 to two of said regions.

- The mycobacterium of claim 14, wherein said deletions comprise a 15. 1 deletion selected from the group consisting of BCGa1, BCGa2, and BCGa3. 2 16. A method of distinguishing between an attenuated and a virulent 1 mycobacterium, said method comprising detecting the presence or absence of a first 2 nucleic acid that hybridizes under stringent conditions with a second nucleic acid or a 3 4 complement of said second nucleic acid where said second nucleic acid or complement of 5 said second nucleic acid is selected from the group consisting of BCGala, BCGalb. BCGA2a, BCGA2b, BCGA3a, BCGA3b, BCGA1ab, BCGA2ab, BCGA3ab, BCGA1. 6 7 BCG<sub>\(\Delta\)2</sub>, and BCG<sub>\(\Delta\)3.</sub> 1 17. The method of claim 16, wherein said first nucleic acid specifically hybridizes under stringent conditions to a nucleic acid from BCG, but not to a nucleic 2 3 acid from Mycobacterium tuberculosis or Mycobacterium bovis, or where said first nucleic acid specifically hybridizes under stringent conditions to a nucleic acid from 4 5 Mycobacterium tuberculosis or Mycobacterium bovis, but not to a nucleic acid from 6 BCG. 1 18. The method of claim 17, wherein said first sequence is amplified 2 prior to detection. 1 19. The method of claim 17, wherein said first sequence is amplified 2 by the polymerase chain reaction. 1 20. A method of claim 17, wherein said detecting comprises a Southern 2 blot.
- 1 21. A method of claim 17, wherein said detecting comprises detecting a polypeptide encoded by said first nucleic acid.

2	22. The method of claim 21, wherein the polypeptide is encoded by an
3	intact open reading frame of a nucleotide sequence selected from the group consisting of
4	BCGa1, BCGa2, and BCGa3.
1	23. The method of claim 21, wherein the polypeptide is visualized by
2	antibody hybridization.
1	24. A method for determining whether an attenuated or a virulent
2	Mycobacterium is present in a sample comprising:
3	providing a first nucleic acid that hybridizes under stringent conditions
4	with a second nucleic acid or a complement of said second nucleic acid where said
5	second nucleic acid or complement of said second nucleic acid is selected from the group
6	consisting of BCGala, BCGalb, BCGa2a, BCGa2b, BCGa3a, BCGa3b, BCGalab,
7	BCG\(\triangle^2\)ab, BCG\(\triangle^3\), BCG\(\triangle^2\), and BCG\(\triangle^3\); and
8	hybridizing said first nucleic acid to the biological sample.
1	25. The method of claim 24, wherein said first nucleic acid specifically
2	hybridizes under stringent conditions to a nucleic acid from BCG, but not to a nucleic
3	acid from Mycobacterium tuberculosis or Mycobacterium bovis, or where said first
4	nucleic acid specifically hybridizes under stringent conditions to a nucleic acid from
5	Mycobacterium tuberculosis or Mycobacterium bovis, but not to a nucleic acid from
6	BCG.
1	26. A method of producing an attenuated Mycobacterium species, said
2	method comprising deleting from the genomic DNA of a virulent mycobacterium a first
3	nucleic acid that specifically hybridizes under stringent conditions with a second nucleic
4	acid or a complement of said second nucleic acid where said second nucleic acid or
5	complement of said second nucleic acid is selected from the group consisting of BCGa1,
5	BCGa2, and BCGa3.

1100 1400 1600 1800 2000 CGATCGCCCC CCCCATTICT CCCCGAACA GTCCCCAGCA CGTCCCGTIC 2100 2300 3000 COTCOCCCC ACCTATTCGC ANATCGACGA COGCCCCCC COCCTCTTCG 3100 3600 AATCCOSTCC 1000 CCCCCCACA 1200 OCCANTOGOT 1300 TCTCCCCCAG 1500 GTCAATGAAA CCCAGATCGA CCGGATTACC CGCGAGATCC CGGCGAATCG 1700 ACCOCRICCTO COGRACOTA TOCACCTOCA CGAACTOGAC COGAACCOC 1900 CTAGACGAGG 2200 OCCCCACATA 2400 CTCCGACAAT 2500 CCCTOCICAG GCATTTICTICG TCTCCCCAGA CCCCAAAGAG 2600 GOTTANGATT ATTICATIOC COCIGIAGCA GGACCCGAGC 2700 AACOCITICI GIACCCCATA CAAATACACG GACCCAACAA 2800 GCTCCACCGC 2900 acceptance 3200 OCCURCAG PICACCOCOC OCCHONACTO TOTOGGAGAA 3300 CCTCCCATCC 3400 AGATCCCCC CAACCACATC ACCCAGCCCG TCCTTACCCC 3500 GATCCCOCC CCACCCAGAG CACCACGAAC CCGATCTTCG 3700 CCCTCGGCCA ACTGGGTGAG ATGAGCGGCC CGATGCAGCA 3800 OCOCOCCTOS TOCOSCOCA GTECCTACET OCCOCACETG 4000 COTOCCCC COACCOCTCG COCACCCG 4200 900 800 900 TOSTCACCO GATOCCCCAG GTCTTGACCG CAGAACTCCA GAGCCCTICTIC COCCOTOCCC GAATACGAGA ACTACCOCGA ACCCOGTOCC GCTGTTGCAG AGTCACCCG ACTTCATCGG GCTGTTCGAC CGGATCTGCC TCACATATCG GCCATACATG CCGCCGGCG CACACGTCGA AACCAATGGT GAAGCCGGAG CCCCCCCAC GACGATOCAA AGCGCAGCGA TGAGGAGGAG CGGCGCGCAT GACTGCTGAA CACGTCATCA TCTCCACGCC ACGCTGGACA GAGCTGAAGT CCARCITICA COCCIOCAC ACCOCCCATA ACCTOCITGA OCCGATCACC CCAACCCCCA TCAGGTGAAT CTCCTGCTCA CCGACTTCAA AGGTGGTTCA ACCTTCCTGG CCCCTTTCTC CGATCGCGAT TCTCCCCCT GACCOCCITIC ACATCGGAGG GCATCCAATT CCTCGCTTCC **GCCGACGAGG AAGCCGCGCA GATGGGCCTG** CATTTCACGT CCGTATICTICS AACCAGGCAG CCCTICOCAAT GGAGGTCTAC OTCATOCCOS COCTOCTOS COGATOSTOS OCGACOSCIG ANCAGACTTIC CCGGCCACCC GGGCCGGAAG ACTTGCCAAC COCOCAGGAG OCAGGTAATT TCGAGCGGAT OCCUCANCOS CICARDOCCIOS CICARDOCCIOCO CTGACGACOG COCAGCTACG CTCCCGTTCG TOGTOGAGCG CAGACCOGCO GTGTTCGCAT CGACAAACTG GAGCCAAACC COCTGACCCG TGAATCOCCT OCCTACAAGA TCTCCCTCCC CCCGTTCACC CTACAGGATG **GCTGATCTAT CTCGAAAACC TTCCACACGT COGTOCOCTA** COCCARCOG MACCACCTT CAAGGAACAC CGAGTGGGCT COCCAGACOG ACCTCACOCC GOCTCACTOC CACATGCACA GATOCCOCCG ATOCCACCOC TOCCCCCTT ATTOCCOCCG ACCATOGACA AGITICOTICOS COCCOCATTO GOSTICOSOCO CACCOGNOCC OCAGINACATC ACCAACAAGG AGAOCOGTGT CCOGGAAGTC GACGCTACTG CAGACGATGG TGATGTCGGC CTITICIDATE ATCGACOGAT GOCCCOGITIT TETIGETICICA COCCOCCOCO TCAACCOCAA CAGCOCGTCG OCCACAACOC TCTOCACOC GTGACGGCCG COGROCTARA TACCOCACOS CTGATGOCCO CTACADACCG CGTCAACACA GGCCAAGACC GOOGCACCOG ITCTCGAOGA GOCOCCGACA CTOCCONTCA TOGATGAACC GCCCGCCAT COTTOCAGOS CCAGTOGCOC COCCARCCCA COCCOCCICT CCCCTACCCT ACCICCIONCOT GARCCCARCC GITCOCCTC Accocccc CCCANGCGC **GCTATAACCG** ACCCAAGTGA **OCM/TOCCAC** TOGACOCTICA CONCONCINC TOCTOCOGG ATTATTTCAT **OCTACCCACA GCACCOCCCC** CACCOCCOCC DOCCCCCTCG COCTCCACCA CCCTANTGAC CCCANOCOCC GTCGATCCTT AAGACCGATG GAATTOCTIGO GCACCOTGAT COTOTOGOTO GTOGOAATGA CTCATCCAGA TOGGCOCCAG TCGATCCTCC GACAGGCCGG GATGAAAGTC GGCGCGGCCG TOTICACCAA CATGGCCGAG GACCTACCCC COCTGCCAAC OCTTTTCGTC GTCGTCGACG AGTTCGCCGA OCGINCABAGE GINCELIGADE GINCEATURICE TACTICACTAC CCAGINOCTIC ANTEGENTIG EGCACCACEA GETETEATER ATTECANGGE GTANTEGGEA AGCACCTTCT ACATCAGTGG OCOCCOCC OCAACATOOG TATTOCOCC CCACCTCAAA CCGAGCCCGA CAAGGTCAAC CGGGTCGTCG CAGAGATGCA AGCCGTCATG CHOCOTCACG ATCCAAGTCA ACCCGTTGCG TCCGATCCAT TICCCCGACC TIGAGGGGA GCITICAAGAT CTGGCCGCCC AGGGCTGGG GICCGACTAC COCACTCOCT COGAGATTCC GATCGGCTTG TITCGSTGCGG CCANATCOGC CANGACGACC ATTGCCCACG AGGAATTCCC ATCCAGTGAG TTCAAGGTCA CANCANGTOT TOCCAGGACC AGOCCATGOC CACGACGCCG OCCCOSTICA ACAGACCACT AGACAAGCCG COCOCATTICA CAGGITTCACC OCAGCGATGA GGAGGAGCGG CGCCAACGGC CCGCGCCGCC OCTEATEDEC COTGATEGGE GACAACCECT GEGATTTGCE CCAACCTICA CITICIATICE ATCEACCTAG GIGGCOCCG OCAGITOTICA TOGANAAGCA CCATCTGATO ATCGGCCTGC COCAGATODO TROCCADORA ACCOMACADO CACOTOCOGO ACTACCOCTIC GGOCCTGCTG GACGCGGTGC CCGACACCCA ACTIGOCOGIC AACTIGAAGA AOCOGITICOC GCCGACCGAC OGACOCTOCO COAGGTTGTO CTOCACCAOC TCOCCACTOC GOCACCAAGA TOGAGITICOG GCITIGGTGAC accordaggia ecoccagaga TGAGCCAGGC TTACAAGGCA GACCATCGGG TITICITITICGG CGATCCCTCC CGACATTCCC COCCCCOAT CACCITCTCCG OCCOCCGANG CCGTCCACCA TGATCACCAT GCTGTGGCAC CTTGCGGCTG CAACGCCGAT GATCGCGTTG ACCGAGATOG TCGCCGCTC GAGNAGCTCG ACCCGATOCC 200000000 CACCTCCCC GATCACCCC **GOCAGAGATG** AAGCCCGCTT CCANTCCOCC CTCCGGCGAC CTGAAAACCC AGATCGACCA GGTGGAGTCG ACGGCAGGTT THOCCCACITY CCACACCCTT GTTGTTCACC CTOCCTOCTC ACCTGATCGA AGTCCAGCAT OCCADOCTOC GNATOGANA OCTTCCGCAC ACTGCCGCTG COCTUGGCA TOGAAGACCC GCTCAAGTTC **GCCACATGCA** CATCATTISTIC ACCTISTICAGA CCCCCTACAT CGAGCCTCCA AATCGAGTTC GGGCAATGCT GGAAAAATG TCACATGATC GTGACCGGGC TOGTTCCCGC **GCTCCACCGT** TCCCACCCAC CCCCCCCATC OCCITOCACTIC GAGCITOCCAC CGACAAGGCG CCATACACCC TTCGCTATCA ACACCATCCC CCOCOGITAA CACOCITITIC TCAACACCOG OCTGACCCAG CCGCTGCAGC AGGTGACGTC CCCCACGCCG CTGATGTCTC actrocroca acaccartos GACGAGGACG ACTOGGACGA GAACCATCCG CICANGGTAAA CACACAAAGT CCACTACCTC GTCGACGATT CCAACCCCC CCCTGGCAGC **OCCCAGNACC** CCCAAGACCA CCCCAACACA CTTGCGGCGG GICCOCTICTIC GTACCOCCAG ATOCTOGOGG GTCATCCAGG CCTAATAGGC GATCCAAACC CCCGAAGTAC COCTCAACGA AGACGTTTCC CACTCACCCC COCCIOENTCC TCCCCCTCCC CCCCCCCTCA COCCACCAGA CCTACTCATC CCCCTCAAGC COTOCITICAG TOCOCTICCA GITCCITTCG TCAGCCCGGT **GTAGGCANAT ACTICACCTICG AATGCATCGG** COCTCCAATG ACCCCACCCC CACCAACTTC CAGGCCCCACA GANTGCCCTC CTCCCCACCA **OCTOSTICAC COCCOCTICC** TCAACAAGAC ATTITIOCCCA 30 ₫ 50 2 5 901 3601 <u>6</u> 90 000 2101 2701 3101 === 501 601 1301 201 501 1091 2801 3201 3301 501 3701 Ξ 8 9 8 문 2 ₽ 300 2401 2901 300 3401 108 500 <u></u> **=** □

9009 7700 5800 5900 6000 6100 9500 900 6400 6500 0099 6700 6800 7000 7100 7200 7300 7400 7500 7600 7800 7900 8000 **B**100 8200 8300 8400 8500 8600 8700 8800 8900 9100 9200 9400 5600 5700 9000 9300 5200 5300 5400 CCTCCCCCCA GCGCAACTCC GGTCGCCGTG CCGAGCGACG ANGCTEAAGA AGCTEAAGCE CCAGAAACCG AAGGCCACGA AGCCGCCCAA AGRICACTICGA TGACGCCGGG GTGGTCGACG GGTCACTIGCT GACTICTGGTG ACCICATTICA **CCTCACGACG** GAAACCCCCGA ACCOGNATAC CTCCACTICC CCACCCCCC TTACCCCCCA CCCCCCTGC TCACCTCCCC COTOTOCOGA **OCCUPATICACC** ATTRICTIONICA **OCCCRAMOCG** COTCAGGATG CORPORCEAG GOCCCOGTICG AGCAACCCGG AGGCCCGACG CAATTTCCCC AACACOGGAT CCCAAACCTA CACCCGAACC COCACCTGCA CCCCACGTAC CACCETETICC CCTCOCCAAG CCCTTCCCC GCACCCCCGG CCTCCACCT OCCITCGACGI CGACCATCGC COCCCCCCT COSTICOCOTT TOST CATCAT CACATCATG CCOCAGAAC CCAATGTCCC AGTTAAAGAC **TOTTIOCTICCT** ATOGOGATOC GOOCGTOCITG GGAAACTICGG CCTAGCTITGT OCTOGCCGTTT DESTINATION STRUCTURE CACABOTTIC COCCOCCOCT ACCOCCOTOC ocrececes ATCATCTOGT OCTCCGCAAT CCGGCTTCAG CAACAAGACC CACCITCTATG GTCAGCCAAG GACGGTGCCG ACCACCATCA AGGCGTTCCC ACCACAACGT CANCARGOGA ACCAGTECET GACEAAGETE OCAGEGGET CAATCCCCG ACCCCCGCC GCTCCGTCCA OCCUPAGGAC ACCACATTG COCCCOGAAC CGAGATTTCA GACGATECTG ACCODEAGAE GGATGACEGA TITTOGTACTE OCCUPATION COORTINGE OCCUPANTE AATATICGIC AGGCCGCGT CCAATACICG COOCMACOCC GAGITCOCGT AGAATAGCGA TCTCAGAGGT TGTCATGGGG GCCGACTACG CAGROCTINCO TITICCOCCGG COCCCGCATC ACCOUNTAGES TOCOCCCCC COCCCACC CCCACCCCCA COCCCCAIC CANACCACC ACACCTCCGN TOCCCATCGC GGAGCGCCGC AGCAACCGGA ATCACCCGCG COCCOCCAG CTCCCCCCG GAACGGAGCC CTCCCCAGCG TCAACCTEGG CCTGTCACCC GACGAGAAGT ACGAGCTGGA CANADCTOCG OCTOOCNANA CCACCCTGAC AGCAGCGTTG GCCUGAAACC TYCCCGATCG GGTACKXCGA CAATYCGCCG CAATGCGGTC AATCTIGGAAG TGCTGCCGGC ICCTOCGICG AGGITITACA ACCICGICIT GGCIGATIGE GACGATTITCG AGAGGCTGG ACGTCGTTGA GCCCACCTGC OCCITICATIVE CUCITATICAL COCTATICACIO GIICATICACIO GITOCTCGAT CCCGTCGCGA CCCCGGAGGC TACCAGCGAA GACCOCAGCA AACTOCCCAA GCAACTTICTC ATCGGATACG TETTITICIACA CAGCETOGTO GTEGEGGGETT TEATEACGAC ACCACACGC GCACCTTTCC ATACCTTTCG CCCCCGACAA ACCAMOCIGE COOCOCATEA CEGOCICAGE AGEREGOEGO ACCEACCAAA AACATGACAG ACCAGCAGTG CAACGCCTG CAGAACCTGG ACCCCCCATG CCCATCGCCG TCACCOCTAT CCCACACACA CCGAACGAAA OCCUTCOTCG CCGCAAGCGT GTCCTCCCAA GTGTCTCAAT CGACGCCGCA CAACAGGCGT COCTTICGACT GTOCTTICACG ACTCACCTICA GTTCGACCGC CCACAGGGG CACCTGGCCG AGTGCCTACT TCACCGTCGC GINGCOCITCO TROTIGOTICOS GITCOATORCI CACOTECEDO OCTOCTOTOS ATCACCOCAS TOTACGACAC CAATOCCTGA ACCOTTCGCC GTCCATCCCA GATTCGGTGG TAGCAGCAAT CCAACATGAA COCOCCOOCG TOCKSTOCK TOCKSTOCKTO CHOSCACCCC GROTHGFINGC CAGACGATCA GTCAAACCGC CCAACAGGCC GCCGGTCCAC GAAGTGACAA CGACTCCCCC ACCCTCAACA COCCOCCION TOTACTICOCIC GOCHETICAT TETCACGAAT GCGCCCAAGC CCCAOCTICAA CAACCCACTG CACCOCAACG COCCCTGACT CGAACAGCAT ACCAMOCOCA OCCOCATCTA AACCACCCAC MODICATION COCCOCTICAG GTCCATCCGT CATCTCCACG ACCAMAGAA CCCACCCCTA ICITICOCATA TTGAGCGCAC TCTTCCACCC MCGCCMCC **GCAAAGATGC** 0000000000 CCCCCAGCC ATTACCOCCG ACACTAGCGT TTCCAACACA COCCATICOCC ACAGGTTCTA **OCCGACCIC** TCATTCCCAT **CCACATCCCC** CCATCCCCGT refreeded TOCOCCAGAC CCCCATCCCT CCCCCACCG DOCTOCCCCC STOCTOCCCT PECCHECIT **AACTCGACGA** CAAGGGGCCCG TTCCCAGGTG TOCTOGODAN TOCCOTTOTO ACCOCITANT CGTCCATTCA **STCACTOOGA** GTTTATACGT **GCGCAGCCGT** ACCTIGITCOCA **OCCOGNACCO** CCACCCCAAC CGACACCACA ACCCCCTGG TCCCGACGTG AGGAAGCATC GACAGAACCT ACCTIGATTICA TTCACCCCAA CONCINCTAL CACATOCCCG TOSTCOSTOT GCATCCAGGC AGCTICTICGCA CTACAACGAC ATCCGCCCAC TCATCGCCGA COGNETICON COCCATOCG OCANGITICAN COCOCOCOCO TOGTOGTICAT CTACANGCOC ANDSTRUCTICO ANTITOSCCOC AGCOCTATICO STREETGEATE SECUCIOSCIO ANOCITORACE ATGTCATCGA CGTCATCGGG TTCCTCCCCA 10000000 COCCOCCCC CTCCCAACCG TCATGAGTTG ACTROGRECE CECEGOGGG ATCCCATTCG ACAACGAGGA CCCCCTCACC GROCCATICOC DOTICOTOCTIC COCODOCACT TIGITICACC TCTACCTCAC GGTAGCCCTG ATCCCTCCCA CAGAACATCG **AACACGTATA** CCCTCAAAA GCCCACCTC **ACCTACCAGE GTGTCCAGCA AAAATGGGAC** CCACCCOOCT CCTTOCTOOC CCTCCCCCCC CCCTCAAACC TOCATTICOCC ACCACCCAGC MANACCCATC CCAGTCCCCT **CCTIGHTSCT** COCATOCTOS CTCTAGACGC CAGCGACGC GACTGCCATT TOCOCCOCCO ACCTOCCACO CCCCACCAGC ccinaccacaa chacaaacid GITACCAAGA TITICOCCAGC GACGACACCG TCGCGGTGCT CTACCCACCG TTGGTCCACG OCCOCCOCC ATCCCCCTTT TCACCCCCC COCHOCOCTO OCCOTOCCOT OCTACCCOOC GALACCOTOS OCCIGATIOG COCITICCTICA GTCACAGGTC ACCUATING CAGNIOCAGG COCUNTOCAG GGANTOTCA TODOTTOGAC CGAAGOCAAC CCTTTCTCGT ANDCTCCGGA CGATATGGCA OCCOCADGNA AGCCOCCTC **GCCCATC**GCC CCCAGACCAC CCCCCAACAC **OCCCCOCCCC GCCCAACATG CCGCGGCGCA** MATCCTTA GOCCOCCCC OCCUPATIO OCTOCATOCO CCACGGTGTC AACCGCTGCG COGCCTGCCA TCGCTGTGCT GGTAGGCAGC COTCCOCCT TOTOCOTOCC AATTIGATOGA COOCOCCATG CTCCATCCCC TITTCCCCAG SOCIOCIC TOACCCAGTA CAGATGTCCC ATOCCCOCAC ACCTIGCCAG CONCERNOCO COCCCAGCGG TOTAGACCIOG OCCITOCCAGO CIXCHOCOTOT COCACOCAGO CCMCCCTCC CCCCCCCCCA cccoccrac **COCCHOCTET** TCGCTGCCCC **COTTCACTICAC** COACCTTCCG 1COCTCTCCC CACCCCCAT **GTTGGCCCCC** CCCCCACCC GTAMATTIG CCCCCTCCT **OCACACCCGT CCACCCCCTT** ACCCCCCCCA ACCOGNOCC CCAACCAGTT TCCAMGAMGC OCADOCCOCTO **GCTCAGGCAA** GTCCATCCAT GCGGGCCGAC CCACALANAG CCCMCMCG ATCCTGGGGA TCCCAACCCC GGATACCAGG TOCCTOCCTO COMPLETE GACCCCCATC **CCCCACCTCC** CCCCCCCAAC COCITICOCAG CANCETATOG GATOCCGATC AACCACCCA CCGNATCCCA **CCCACATCAN** GANCCACCGA GTATCTOOCT CCATTTAGCC OCCACACAAG CCAGTCCGCC GCAATCCCCG **ACTICACCCOC** ATTTCCAACA 2000000000 **accritecene** OCACCEMOCG CTTCATCACG COCCGATTCC CATCATCGCG CTCCCTCCCC **CCCANATTICS** TOCAATCOCT CACCTOCTICA CHOCOCOCC TOGATACCOG ACCIOCCAGAC CAGCGCGCCT ACCOCCCCT TOCCGATOGA AACTTATATT CCCTTTACCC **COCCURCO** TCAGTCACTG ACCIDENTICE **AACCGAGATC** CTTCCCCTAT GTOCTOCOCT ACCACCACCA OCTATICGAGG COCCOCTAG CACCCAAGCC COCCCACTT CCGGCCCGCAC CCGAAGCCCA COCCAACTCC COCCCCACCC CCCACCCCAA CCTCCCACCC GICCCCCCCCC GTACCAACTG COTCCGACCC CATCCTCGAC ACTOCATOTO TOOCTICAGGE TCATCTCCTT ACCTCCCCCC TCTTCCACCC **OCACTOSTTG** CTOCTOCOCC TOGACCCTAT **GCTCCTACCG** ACCCITCTOG TCACTCACTC ATCCCTTTICT **GGCGATTTGGC** TATCTOCTGA TGTTTTTCAC ATCCCCCTCC CCTCCCACCC CACCCTGATT TITICOCTICOC ACCCCACTA OCTCCTAAAA ATCCGGTTCT TTCACCCTCT COCHOCOCC ATCCCAACCA CCAAGACCCA TCAGGCTACC CAACTCGTTC CCCAGGGCCCC COCCACCAG **OCCEANDOOR** 6401 1017 **B**101 8201 8301 1096 5201 5501 6601 90 7301 7401 7501 1601 107 1801 9001 9301 100 1901 5001 5101 \$401 5601 5701 1085 901 6001 101 6201 5501 6901 1901 3001 1401 1501 8601 3701 8801 1066 9101 201 <u>8</u> 80 5301 6301 5801 9401 501

12200 12400 12800 13000 13100 13200 13300 13400 13500 13600 13700 13800 1 3900 14000 14100 14200 14300 1400 14500 14600 12300 12500 12600 12700 12900 11500 11600 11800 12000 12100 GICCTIATAC GUGGACGCG COOCTGCGC 11700 CHICHTCATC TGATTGAGCG TCCCSTCTAG GGCCGACTGA AACCGCCCG 11900 10700 10800 10900 11000 11100 11200 11300 11400 10500 10600 10300 10400 GCACTIOCCS COCCCTITGTC COTGATGCGC GCCACCCATC COCATTCCCA CCATCCOCTA COCCICANO TCGCTCCGC TGGCATCCCG GTACCGGGAG ACCOCCOSTG OCCACCCTCG CCSTATCGGT TAGTTCGGCC GAACTGTCCC TETTICAGENA ACCATETTEG ATGCGCAGGE COGTGAGCCA CTGGTGCCCA COOCCUATION CACCAGCISC ACCOCCICC ACCOATICGS COCCCCCAT TOCTICTIOCG GAGGGGGG GGGTCGATCT TGATGCCGGG GGGGAGGCTT ATAAGTCOGA TOTTCCCGCC TAGCCCACAC GTGCAGCTGC GCGACATATT **CCCTTOGTCG** COTATTCCCG CATOTTCTCG GCGGACAATA CCAGCTGTTG GGCGCGTTT TATACCCATT ATCGTCGCTA AACTGAAAGG TTCCTGCACT AATTTGATGC TACCOCCATG COCCCGAGA TAGGTCUTCG CATTCOCCCA CGCCACCTTT CACCAATCCG CCGATCCATG GTCCCTAGTC CCTACCACCG CAACAATGCG **OCCURRENT** OCCUPACION DE L'ANTIGNOCION ACADOCION CONTINUES COMPENSION CONTINUENTA TACTOCOGOST ACACOGOGAG AATTTOGTOG CATCCCCTTC TECACECES TECEAGEETE CTUATCAACE TECECATAGE CETTICACEC **TCTTICCTGAT** TCCCCACCTC ACCACGAACT CCAGTACCCC **OCCCCTOGGT** CCCCCCCACC TOTOGOTOGO ATACCOOCC **acceccact** ACTCCCCACA TCGGCCAATG **OCTOCTICAAC** ACCCCACACC ACTOCCCTCC TTCCCAGTAA CCCTCCTAAC CCCCTCATCG GCACCCCATC CAATATOCCT **OCCCOGTICAT** ATTICITIOGAL TCACCATCCA **OCCOCCOGAC** OCTUADOCCOC COCCOADOGG TOGACGACCT ATCCAACCGT ACCONCATOR TENANGICIT CCACCCATCC GCANAGTUCG CGAGCGATOC COCCATCOCO GCCTCCACCT CGTTGGCCCT GTTCAAAATC ATCTIGGTAGC COCTITICCTC GOGTGGGGAA ACCCGGCGAA AAGGCATGT AGACCGGGCA TCGGTTCACC GTCTCGCCGA ACAAAGCCGA TOCCOGTOTT GATCGAGCGA AGCCTCGCAA GCGGTAGCCG CCCCCAAATA GCCGCCGAA GTGGTCCGCG ACCCCACCA CCCCGACCAC AGGTTCTCAC OCCUMENCE GAGGEAAGCA AGGESTGCCC COTANGGCG ACCOCCACCA CCCCCCACCC TECAGGIECE GEOGITEGEE TOCOCCIOCA **GCCTCAATGG** GCCACCATCA COCCACGACG GAGGTTTCCA TCATCCCCCT ATCCCCCTCA OCATCCACCC COCOCCOUTA GCCCTGTCGC ACCGGGCGCT CCGAACCCGG PEGACTECETT ACTIGITECTEG COCCGACGGT TACCAATGAC PACTOGREAT GOSCHICATE AGETEGAACE ACAGEATOTO GNACTOCCCA COCOCCAACC COCCGACACC GOGATCATCG ACACCGOCGT concentent ATCAACCCGG ATATTCCTOC CTICATCCCCT ACOTOMOTICO CAGNOTOCCOC TITOCHOCAGO GCAATCTCCT AACCOGGTAG GTCGCACAGC GTGCAATFTC CATTGGCGGC CACCTGCCGC TOCTICATOCT THICTCCCC COCCUCACO TCACACCATA TOTOCOCTTG CCCOTCGACG ATGACCGTCG COCCUTAGNA AGTICACCOTIG TAGOCCAGC CATAGCTGTT TCGGACGCGT TGAGCGCCCGC CGCGATGCGT NAME TO CAST COUNTY OF ACCTROCTICA CCCACCCCC CCCTTCCCAG AMCCTGGCT CCTGGTAATA ATUCCGCCCT ACCOGGATCC TOCCCCCACT GTGACCATCG COCCCCTTGA TCATCGACCA TTCACGGTGG COCCOCTOC CCCCOCCCC ACCCOGACTC TACCCACTCG ATCCCCAATG COCHOCACGA **OCTACCCOC** CCACCCCAA TOCOCTOOCC ACCCAGGGCG ACCAATGAAT ACCCTITICCA TOCCCCTCGG CTCCCCCCCT GACCAGGACT COCTICATING 200000000 TCCAGCACAA TCTTGGCCAC COCCULCOCC TATCOCATTT **GIGTCOCCA** TTOCCGACGC ATCTCCTOGC TOCATOOCOG CTTCCACCCC COCCICCOCC CTCATCCOCC **CCCCCAGGC** CTGATCGGTA GCACCICCAAA GTCCCCTATG COCNERSIAN corcoccina TTATTGCCGA TACATCOCTG CTCCCATAGC TCACCTCOCT CANTATTOCC COCATTOCOG GICACCTOCT CCCCCCAGC COCACCACCT TCATCTTTOC TTCACCCCCC OCCGATICAAC COOCCAAATT CCTCGAACTG COCCACCCC TCACATTGTG CCGCAGAATC ACCCGGTCAA GTAGGCATCG COCTOCOGTA CCACCAGTCC GOTGTCCAGC ATGTCGTCGG CCCCAACGTG ACACCOCCOC CACCCAGCCG COTCTCCANG ANTOCCATCA CATOCACCOC TYGGGCGATC AGCCGGGTGG TECTTEGITE ATCCCCCCC ACATCCCCA TAACCCACC ATTIGTISCADE SCENESTIGA CECUCITAÃO COCODECTES GEACCEAGET CTCGGCTTCG TCGGTGGCGC GGAAGGATCC CCTCGGTGTA GACCCCTCG TCTTCCTGCT GAGAACCCTT OCOCCOCCO COOCCCTICC CTOOCCTAAG CACCACCCGG GATCAATCCC TCCAGCCCCA CTATGTCTTC GAGTATCCAT ACCOCTTGAC GTCGCCTTGC AGCGTCAGGT AGTTCCCCTTC GACCITICCO CITCITITICOCA CCOCCOCCA TCCTTAGTGC GACATCGTCG AGATGCTGGC **CCCCCCAT** TCATCGAAGC TOCGGCCACA CACCGCGTCG ACACCATGGC COCADECETA AAGAITTET CATEGOEGOE TEAGEGEEEG TETAGEEADE TOCCCCATG **GCCATCCACA** ATCAATTCAT AGCITGTCCAA CACCATCTGT CCTCCCCCAT CTACCTCCCC **GCTGAATGAC** CATGCTCACT CCCCCACCGA TCGCTCAGCA **accencents** recensoon COCTOCOTTO CCGGTAGCGG CGAGCAAACG GICATCTOOC COCCITICATIO **acercacens** ACCIONACCION COCCACACG CTOGACGCC APPECATECE ATOCOGACCC **GCCCAGGCTT** COCCOCCICCA COCCENETCA OCCOGCAGOC accocates acadaceas TOCCATICTIC AGATCCATGA GOGACATCAG TOCCICATAT COGATCATCC COCCCTACTIT COCCUATICAL ATCOCCACCT COCAGTICCC CITICOCCACC ACCCITCGACT AGCCGCCAGT **CCCGATTTCC** CACCTOCTCA TACAGACTCA CCACGTCCGG TOCCAOCOCA CCMATTCGTC GPTCFTCTAT CAGITICCTTC GTAACCGAAC CTAAGACCAG COCCTCCCCC COCCOCCOC OCCANCAGE CAGTOCCAGT CTOCAOCGTC GROCCITTIC TOCTODCOOC тесифсиос соссеосите GTCCGGTTGG TCTCCCGAAT STOGAMOCC TTATTCOOCA COTAGTICCCC TOCAGTOGCO TCATCCCTCC CICCOCCACC ACACTGACAT CCCCTTCACC ACTICACOCCIG TIGTICACICIC ACCACGCGT ACCAGTAGOG CTGAGCCTCA CCCAGGTTGA CCAGCGTCAA CGCCCGGGCG ACATGCGGGT **GCGCTCAATT** COCTICATIO OCTCAGCAAC COGACCTICCG GOCGCCGCGT **CCCCCCTCC** COCCOACCAG AAACTCOCCT ACCOCCOCCA COCTOOCOOD COCTOOCAGC ACCIOCACCCC **COOCTACCTT** GTANAGCCGT COCEMENTS TCTCCCTCCA CCCAACCCAT CANGTACACC TITICICCATG STECCAGGIA ATCACCGIGG CGAGATAATC CAGICCOCCA **OCHENTENDE** ATTITIOCCOC **CCGTCATIGTC** CHOMOGRACC TOCOTOCOCO ATATCGATCT OCCUPATION CCCCTTCATC 200200000 COCCTACCGA **NOGNOCTICCT** DOCCADOSCA ATROCORDOCC MCCCCCC OCCATCOCA ACCCCCCGCA TOAGTTCGCA **GETICTICACGAC** TCAAAACATT GACAATGCGT OCCCANOCACO TCACCTCCCC ACCOGNICCA **OCTUTIVOCOC** CCCCCCCAT CACCTCCCCA TOCCOOCATC CCATGTGTGC TOTOCOCCGT CTTCCACCGT **OCCUPICAGAC CONTROCTING** CCAGTACAGE TGACATCCAC CCOCCCCTC ACCCCCCCCA COCTOCTICAC COCOGTGATC TOCCATCOOD CCTTOTCCGA GGTAGCCACC ATTICTCCAGG GGACCITCATA ATTAACCAGC COCMACOCCC CCCCCTGCC CCTCACCCGT CTTCGAAGAC TTCATCCCCT AGATECOCCTC CCCCCCCAT ACCTCCATTA 0000000000 GROCITOCIAG ACTCCTGCGA CTGTCCCCAT CCAAACCGCC CTTAGGCCAT TTCATCGTCA CATGCCCGAT CCCCCATCCC CCACCCCACG TCACACCCCA COOCOGETTE CCAGTTGTCA CCCGTTTCCA CTCCCACCOC CTTCCCCACC TTACCCCCC CCACCGTCAC COCCCGTTCA CCGATGCCGC ACCTICCTIGGE COCTCAATGT ACCTGACCCG GICTIOGGCT ACCATACATO GTOCACCGAC OCTACCGATC TOCCACCGCC COCCCCANG CACCTOCTT CCACCCCCCA CCCCTTGAAT ATTOCGIGAT TCCAGCACCC COCTACACTC AGAMGTGTC TCACCTATTG **NOCUTAGOCC** CCCGGTGGAT **OCHGACTCGC** CATAGGCCGC 13201 13901 13601 14101 11901 12101 12201 12301 12401 12501 12601 12701 13801 3001 13101 13301 13401 13501 13701 13601 13901 1007 4201 1301 500 4501 4601 1201 11601 12001 10601 11201 1301 11401 11501 11701 11801 10101 10901 10801 1001 11101 10201 10401 10501 10701

	15100	15200	5300	5400	5500	15600	15700	15800	15900	16000	16100	16200	16300	16400	00591	00991	00/91	16800	58891	
_		_ `	_	_	_	_					_	_	_		_		_		16	100
	ATCTGCGCCG	CCTCCCGCC	CCACC	cerocerce	CCCCATAGIC	<b>CCCCCCCCTCC</b>	<b>GCCTCAGACG</b>	TCCCA	GICGATCACC	CCCOCCAGCG	OCCITCATOCC	TOCCATTICAT	2000	GATCCCCXXC	CCCCCACGTT	CCTCCCCACT	SCIENCE	<b>GTACCCCCCC</b>		_
	TG ATC	ည	es es				AG GCC	GA GAC	स स	אכ ככנ	300	AT TC	युट ध्र		8	5	<b>1</b> 8€	CG OF		90
	CCGCTCCATG	OCCOGCGATG OCTACCOCCC	CCTOCATCGA GAACGAGCTG	<b>GTTCCAGCCG</b>	TICACGITICA	ATCACACCCA CGCCGAGGTT	COCTTCCAAG	COCCATIGATO GATOCTISTICA GAGTICCCATIG	CCACACCCCT	GCCGATCGTG GAACCCCGGAC	ACCCCGATCG ATCCCCCGTG	AGATTC	7000	CCTACCAAAC	CCTGAACCGA TACCGGTTGG	TICCATAGCG TCGATACCGT	CTOCCOCCC GTGTAGCGTA OCOGTGTCCA	GACGAATTICG	ACTITIC	_
	אד כמ	20 02				SC CG	AC CO	ITG CAT	05 55	TC CA	CG ATK	or To	SAC OCC	ट्र	CA TAC	xc Ta	XG GT			80
	TOCCCCTCAT	වි දරවට	ACCCACOCTC	TCTGCACATT	CACCACACACT	CACACC	GTGAGCCGAC	CCATC	CCCCAACCGT	CCATC	CCGGA1	CCATTC	COTCIE	ATATCCCCCT	TCAACC	GCATAC	20000	COTTCACGGT	TCACCTAAAT	_
,	500	8 5	00 AC	56		TG AT		22 22		800		5	ACA AC	CCG AT					ore are	20
	TOTOTOGOGG	CAMCCACCT	<b>COCCACOTOC</b>	GTGACAACGG	CCACCACCAC	ACTGATGITG	остост	2220000000	COCCACCCCT	TTOCTCCACG	GCCCCAACCC	TECEAGNACA CTECATTEGT TEAGATTEAT	TOCOCAMACA ACCOTOTOR OCCTOCOCTO GTOCOCOTOG	<b>CCACTACCCG</b>	ATCTCCAGAT	ACCGACCGTT	ATCCAGCTOG	<b>GCCANACCCG</b>	TTCTATAGTG	_
;			8	5 5 5 5 7				800				CAC TC	200		ATG AT	CGA AC	CTC A1	8 5	GTA TH	09
	00000000	TATCCCCCCC	CTOCOGOCOC	CCACCCCCCCC	MCCCOCCG	ACCCCCTTC	DOGGICATIC	TCGGTCGCC	CACCITICCCC	CAGATACGTG	<b>OCCACGTCGG</b>	TACOGTOCAC	ACCTAACCCC	<b>ITTTOCTACC</b>	GOCCITICATIC	TTGCGGGCGA	GTOCOOCGTC	GAACCOCCCA	AGCTTCAGTA	_
		٠.	E		2	_							Ī	8		•		_		20
Š	દુ	Ç	E	g	Ø	E	ರ್ಷ	۲	ŭ	Š	ŏ	3	F	¥	ဗ	5	٤	8	8	
3	CCTCCACC	ACCCGACACA	ACCATCITIT	<b>acastacc</b>	CCACCA	ACCTITICTAGE	COGTOCC	GACGAACC	AAATCCCC	COCCCACC	ATCCGGG	ATCAGAM	TCCATTT	GACTICAAC	CGATGTCG	TTCGATCACG	AATACCTO	actimicacti	TTOCATOC	
	CAGGT COTCGACCCC			CACCA OCCOSTOCCTA	CANTE OCCACCOSTIG		DOCCE COGREGORY	ACCCC GACGAACCCG	CGACG ANATCGCC	OCCAT COCCACOCC	DOCTIG ATTCCGGGCCC	COOTIC ATCAGAAGA	CHICTIC TICCATTITITAG	OCCOG GACTICAACOG	SCCAT CGATGTCGGT		CCACC AATACCTCCC			40
CALLACTOR COLONAC	CCCACCAGGT	COCCOCCAG	OCCUPACIOCO	CTTAACACCA	GTACGCAATC	COCCTOCTC	CCCCCCCCCCC	OCCICIANCOC	CTOCTCGACG AAATCGCCGC	THOTOGOCAT COCCCACO	COGTAGOCTG ATCCGGG	COCCACCGTG	COCACCTICTIC TOCATITITY	ACCACGCCGG GACTICAAC	ATCOCCCCAT CGATGTCG	CTTCGGCGC TTCGATCA	COCCITCCACC	TOCANCOCCA	TOCOMOGG	- 04
CACIC CACLACTOR COLORAGO	CCCACCAGGT	COCCOCCAG	OCCUPACIOCO	CTTAACACCA	GTACGCAATC	COCCTOCTC	CCCCCCCCCCC	OCCICIANCOC		THETCOCCAT	COGTAGOCTG	COCCACCGTG	COCACCTIGTG	ACCACGCCGG	ATCCCCCCAT	CTTCOCCCC	COCCITCCACC	TOCANCOCCA	TOCOMOGG	30   40
CALLEGET CALLACTOR WINNING	ACCECCACCO CCCACCAGGT	SCCCANGCA GOOCCOCCAG	ACCESSIFIC OCCEANDOORS	CCCCCGACG CTTAACACCA	COCCATTOT GTACGCAATC	PTCATCGAT COOCCTGCTC	COGATOGAA CCOOCOOCCO	SACCITICAS OCCICIANCOCC	STTACCOCC	STUGGERAC THETGOGGAT	DECOCTIST COSTAGOCTS	CHACGCCAG COCCACCTG	ACCACTTCA COCACCTGTG	IGTACCTOG ACCACGCCGG	ACCCTGCGT ATCCCGCCAT	DECEMBER CTTEGGEGG	SCHOOLIGAC COCCITICAGE	SANTAGAC TCGAACCGCA	COGNODO TOCCANCOCO	° -
CATEGO CACEECUCITE CACEACTUMO UNIVERME	ACCECCACCO CCCACCAGGT	SCCCANGCA GOOCCOCCAG	ACCESSIFIC OCCEANDOORS	CCCCCGACG CTTAACACCA	COCCATTOT GTACGCAATC	PTCATCGAT COOCCTGCTC	COGATOGAA CCOOCOOCCO	SACCITICAS OCCICIANCOCC	STTACCOCC	STUGGERAC THETGOGGAT	DECOCTIST COSTAGOCTS	CHACGCCAG COCCACCTG	ACCACTTCA COCACCTGTG	IGTACCTOG ACCACGCCGG	ACCCTGCGT ATCCCGCCAT	DECEMBER CTTEGGEGG	SCHOOLIGAC COCCITICAGE	SANTAGAC TCGAACCGCA	COGNODO TOCCANCOCO	° -
S CENCENTUS CACECUSCIC CALLACTUS UNIVERSAL	ACCECCACCO CCCACCAGGT	SCCCANGCA GOOCCOCCAG	ACCESSIFIC OCCEANDOORS	CCCCCGACG CTTAACACCA	COCCATTOT GTACGCAATC	PTCATCGAT COOCCTGCTC	COGATOGAA CCOOCOOCCO	SACCITICAS OCCICIANCOCC	STTACCOCC	STUGGERAC THETGOGGAT	DECOCTIST COSTAGOCTS	CHACGCCAG COCCACCTG	ACCACTTCA COCACCTGTG	IGTACCTOG ACCACGCCGG	ACCCTGCGT ATCCCGCCAT	DECEMBER CTTEGGEGG	SCHOOLIGAC COCCITICAGE	SANTAGAC TCGAACCGCA	COGNODO TOCCANCOCO	° -
MICIOS COMCINIOS CACICORAIS CALCARONA CARANA	ACCECCACCO CCCACCAGGT	SCCCANGCA GOOCCOCCAG	ACCESSIFIC OCCEANDOORS	CCCCCGACG CTTAACACCA	COCCATTOT GTACGCAATC	PTCATCGAT COOCCTGCTC	COGATOGAA CCOOCOOCCO	SACCITICAS OCCICIANCOCC	STTACCOCC	STUGGERAC THETGOGGAT	DECOCTIST COSTAGOCTS	CHACGCCAG COCCACCTG	ACCACTTCA COCACCTGTG	IGTACCTOG ACCACGCCGG	ACCCTGCGT ATCCCGCCAT	DECEMBER CTTEGGEGG	SCHOOLIGAC COCCITICAGE	SANTAGAC TCGAACCGCA	COGNODO TOCCANCOCO	° -
I CETAMICIOS CEACGATOS GACCOSCIC CACALCASE GALGARAILE	ACCECCACCO CCCACCAGGT	SCCCANGCA GOOCCOCCAG	ACCESSIFIC OCCEANDOORS	CCCCCGACG CTTAACACCA	COCCATTOT GTACGCAATC	PTCATCGAT COOCCTGCTC	COGATOGAA CCOOCOOCCO	SACCITICAS OCCICIANCOCC	STTACCOCC	STUGGERAC THETGOGGAT	DECOCTIST COSTAGOCTS	CHACGCCAG COCCACCTG	ACCACTTCA COCACCTGTG	IGTACCTOG ACCACGCCGG	ACCCTGCGT ATCCCGCCAT	DECEMBER CTTEGGEGG	SCHOOLIGAC COCCITICAGE	SANTAGAC TCGAACCGCA	COGNODO TOCCANCOCO	° -
14301 CEINAILEGO CONCOVILOS QUELLOS CALCACIOS CALCACIOS	ACCECCACCO CCCACCAGGT	SCCCANGCA GOOCCOCCAG	ACCESSIFIC OCCEANDOORS	CCCCCGACG CTTAACACCA	COCCATTOT GTACGCAATC	PTCATCGAT COOCCTGCTC	COGATOGAA CCOOCOOCCO	SACCITICAS OCCICIANCOCC	STTACCOCC	STUGGERAC THETGOGGAT	DECOCTIST COSTAGOCTS	COCCACCGTG	ACCACTTCA COCACCTGTG	IGTACCTOG ACCACGCCGG	ACCCTGCGT ATCCCGCCAT	DECEMBER CTTEGGEGG	SCHOOLIGAC COCCITICAGE	SANTAGAC TCGAACCGCA	COGNODO TOCCANCOCO	° -

Figure 2

Page 1 of

6800 7300 5300 5400 5500 5600 5700 5800 5900 0009 6100 6200 6300 6400 6500 0099 6700 0069 7000 7100 7200 7400 7500 7600 7700 7800 7900 8000 8100 8200 8300 8400 8500 5100 0098 8700 8800 8900 9000 9100 GACCACAGAG ACCTICGACCA ATCCTTCCTC MCACCAGGA CCCCCCAACA ATYXCACAAGA AACGACTICC COCAAACCGA CCTTICACITIC CGCGGCGTCG TCGGCTACTIG CCCCCCTTCC **CCTOSTICGCT** CTCCCCTAAG CCCCCAAACT GAAGCCCACG **CCAGTGCGAC** CTTCCATGCG CAAAAGCCAT ACGCCTCCCT CGACGTTCGG COCOTOCACC **AAGCTCCGTC** CCCCTCTCCC CAACCCCTCT TGCTGGACCC CAATTCAAGG OCCGACGACG 7000000 CCCCCTCAAC TCCCCTTTT CTTGTGTGCT ACCCOGTATT GITTICICITIC CCACCATCGA GTGCCTCGAT GTCCACGCCG CCCTIGITAACC COTOTOCCACC CCCCCTVSCC CONCINCOCC CCCGATCCCG CCCCCCCCAA CTCGTTCCCC GACCCCTIGA TAGGTAGATA TOCGTACCGA CCACACTCTA **CCAACTICTGA CCCCCCCCAT** CCCTCCCCCT CACACCAGCA ACCCCAATGA CCGTCATGTC COCCTTCACC CTOGAATCTG AAGGCCAAGA AACTCCCCAG ACTITICICCC TTCCCCTCTA GAACACGAGC TCACCTGCGG TOCCCAACCC CCACCTTGCC GCGGCACGCT TCTGCCGTCG CCTCGACKAT TCTCCACCAG **ACACCCACC** TCACCAGGAC GAACCCCCC CACACCGTGC AGCAGGCCAG TTATTAATGT COTOCOTCACC ATCOCTOCCA ACCTCCTGCT CCCCCCTCC CCCCCCCTA ATCCTTCCTA COCCAATTAC COACACOGTG AGGCCGCCG CICCCOCCT CCCTCTTTCC CITCCCACCG CTCCCTCCCC CTOCAGATGA CTCCACGCAG CCTTGTTCGC CCCCCCTT CAGCTGCCCG CICCCCCCTC COCTCCCCTG TATACCTCTA AGACCATAAA GTATISTICAG TOGOGGATGA CAAGCCAGGC GCCCACATTIC COTAGAGECG ATCACEGECG GOCCTGGTGT AGACCTCAAT TCGATGATCG ACCCCACCCC TATCCCACTC CCAGAACCCC GAAGATGTAG TACAGCCCCC ACAACAAGAT CAGCGGATCC GACTCCATCA CCGAGGCCGC COCOCCATO CTCCACGCCC CCCCCGCCCC TOCCGATCAA **GCCATTCGGA** TOCOCOUNT CCTCCCCCCT CGAGGCGGCA CAGCACGCGT TCACCGGCTG CTCGACATCT **GCACCGANT** GACACGCCGC CACTCTCACC CCCCATACCC TCTCAACTTC ACCOCACTICA CTICTICACCOC GCCGAGGCCG ACCCCATTAG COCCUATICAG CACGATIGTICG ATCAGCGTGA ATACCTAATA ATTICTTICATT TACATTAGGA **OCCUPATION GCTCGCTGAA OCTETE COCC** CCAACCCCCT AGACCGACGC CACCOCCCT TOGICTCGAT TCCGGACACC AATTTTGACTC COCCICCOCO ACCIACATAA TCACCICACC TCACCTCCCC COGATOCOTIC TTCCGACATC CCCCCCTCCA **GCACCGTCGC** CACCOCATTG GACCCAACTC GIOCOTTOAG GAACOTCAAT OCCOCACAGG CACCAGGICG GCCAGGTGCT CACTGCCGCC OCATICTICCO TTCCCGATOG **GCCAGATCAT** CCCCCCCAT CACCCCCCC CAGATCGTCC MAGCGCICG CCCCCTCCT GACCAGAAAT CCCCACGCCA OCCACCOTICA GOCACAACTIT תמכשבכנטב בכדושבננשב CCGTTGCGCG CACGATCTTC GAACGACTTC ANGGOCOCC ACCOCOCOCOC CCCAACTCCC CCCCCAACCG CATCATCACC GCGATCAGGC PETECECECE APPROCESCO COCCATICACC CCOCCCCCA NOTNOCCCCT AGTCCCAGGC CCCATTOCCT GOCCOCOGTG ACCACCGAGC CCCACCOCTT OCTACCGITIC **GCANGTTOOC** TCTGCACGAA CTGTTGATGG COTOTTCCAC CACATICCAGC coccoccet TGTACACCOC OCAGGTAGCG CGTTGGCACC TACCTCCAAA TATACCITOCC CTCAGGTGAG ATTATCTCCA GATCCACACC COCTRCCCCC TACACCACC OCTOCATTICA CCGCCCCAG ATCAGCGCCA GCGGCCGCAA COCCMACCAC TCTCGGATTT TCAGTAGCTA TCACCTAGAC CCACTAGCCG COCCTACGAC CCCATACCCG CACAGCTGGC GTCCTTCCAG OCCITIOCINCO CONTINUESCO COCCATCTAT CCTTTCGTCG GAACTOOCA CCGATCATCG GITCHICHGT CCCACCTCCC ATCCCCTAAG ATCTCACCAT **CCCCCCATC** COCTTACCTC TOCCAGCCCC PCCCCCATCC TCCCCACAT CTGTAGTCGA CCCACCOGAA **GOCCHOSTICC** GOCCCTCAAC ACCOTCOTOG CCCCCACATC TGTGACATGA CACCACTCAT TOCOCCCC **OCCUPACIO** TOCOCOCCO CCCACCCTOC CCCCTCCCAC GTCAAGAACT OCCATACGC MCANATCCT **GCCCANCCAT** GCAACAGGTC OCCACGCCCG OCCUMOCOC AGCCCATCAA COCCOMMCA ATCCCCACAT **GTGTGGCGAT** ATTCGTCCAG **STOCKTANDO** TOTTCCCCCA ACCCCACCAT CGCTTCAACC **GTTCCCATGA** ACCCCTCCCC GTGTTGAACG CTAGCAACAC GACGGTGTCG AGGTAGACGT ACCAATGCCT ATOCCITCAGE CHICACOCTC AATOCTCCAA ACTCCACAGT CATCAATCCC COCCOOCAGT cocconcer TTCGTCCACT THICTCCACC occoencer CCCATGITTIC TTCCCGTTAC CCAATCCTCC AGACTCCGCTC GACGTTCCCC CCCCCACCAA CCCCAAGTCC CCCAGGATTC CCGTCGGCTC TCCCCCCCC CCCCCCCCTC ccoccocca AACCGTTOCC corrector CAGTTCACCG CCGATACCTC TOCCOCOGAC **GCCTTTTGACA** TCAAAGACCC GITTITICCAG CTCGACCCC OCOCCOCOCOC GACCTITICT CHOOOCCACC MACCACTECT COCCOTCOCC ACCACATTC CCOCCICCGA ACCCGAGGGT **GGATCCGCAG actorcy** CGATCAATGC CTCCCCCACC OCCUENTICE COCCCATTCC CTACCCCACC CCAGCAGGC TOCOGTICOC **OGACATICITIC GCATCCCCCA** COGTICACCO CAGACCICAAA THOOCCTIVIC CCACCACCAC CCACCTCGGT **OCCIO**CACCCCC COCTGAGGCG **GCCTCAAACG** COSTCATCCA GTICTOGGGC MATATTATOG TIGICCCTAT GTAATCGTTT GACCICCIATOC CCACCACCTC COCCATICGIC GFTGTCGGAT CGATCCCATT GTCAAGCACT TTCCCATGTA ACCOCCCCTT CCCCCCCCCC GTCCCCCAAT **GCCCCAACAA** TCCTGACCCC **CCACCACCCC** TTCAAGCCCA CTTGACCGCC COCCCCCCCC CCACATCCGC CATCACCAGG TOCAGACGCT **GCCCTCAATA** CCCAGCTGGC **CCCACCCCAAC OCTCACOGIC** TOCCATCACC CCATGTGCCC ATCCACACTA ATCCCCCTGT TTGCGGTAAA COCCOCCIO CTACCAGTOG ATCCCCTCCA COCCOOCIO CTCTATTGGC OCCOCCCGACC GACCOCATCO COCTTICACOC OCCITOCICA CCCCCCCAOC CAGGGTGCGA CCCAGGCCCC coercocce CATGACATAT ATCCCCCATC CACCCOSTA COCCOCCICA **CCCTTCACC** TACAACAGGC GTCCTGTGGC ACCUITICCC **ACTOGATTICC** CTCACATOCT CACCCCCTCG CITAACGACT COCCCACCG TTGCCGACGT TCTCCACCAT TCAAAGACTG GAACCOCCETT CCACCCCAT CCCCCACCCC ATTAGCGGAC CCCCATACCA MACAGCGCT ATACAGCGTA **accendicas** Accectoccc OCATOCCOCT CCTATIGHTCT TCTACCTGTG GTOCCCTOC **OCCITICATITIST** OCCOMPCCO TOCCCCACA TOCTIGITACTA **CCCAGCGCGC** CAGCCCAAAA CACCCACCC ACCTOCTCAC ACTICACCCA TOSTICTACAC GACCOCAGGT COCATCACCA AGGACCATTC ANTOCOCTAC ANTCOTOACG CCCCACCCC CATGTCTCAG TCCAGTGGTT AATCCOCTTG TCCCTCCCC **OCCTACATGE** CCACCTCCCT CAGCAACGCC ACAGGTTGAA CCACCCCCCT GTCACTCCCC GACGACCTTO **ACTITICITISCI CCCACCACTA** TOCCTATTCG CANTATOCAC CCCCCATGA CATCAAATCA TTCCGGATCC CACCOSTICCTIC GACCACTCCC TGGTTGAGGT **AATAGGCTTIC AACCOSTICETIC** ACGTCCCCTCC GTATCCCCCA OCAOCCAGCC COCCOCCCAC OCCUCCOCOCO COUNTRIES DE COMPONIO COUNTRIES DE COUNTRIES CGAACCTOCC CCCTGCGCCC CAGTGTGGAA **ACTIGISTACCA** CATCCOOCAT CTCCTTCCCC AACCCCCTG STCCCCCAAT TCACCCCCCA CACCCCCCAG TOTOCACACG TGAACAGCCC Gracrocorrio CCAACCCCCC TCACCACATT CACCOGCAGC ACCAGTOCTC OCTOCOCAGA OCCOCCOCNOC GICCOCCCAC GAMCOCACO ATCCMOCACC COCTOOCTCA ACTOOCGTOO OCCUTICACCO GTACCOCCC **OCTTOCTICAL** TCGTGGTTGC CACCCATICAG TACCCAOCAA TIGICATIOG COCCAOCCA CCCCACGATC CCCCCACCC **ACANTOCITIC** ACATACACCT TCACCTCCCA COCCUTT CATCAGGCCG COCTOGATICA TCTTTACGTA ACCICACCCA CCACCACATT OCTIGATOCTIC CGATCACCAT TGACGTACGC TCACCCACCT CACTGACCCC GITCGCTCCC CCCTTCCCCA CCCCCCCCC TCCCGGGTGC CCCCATCCTC TATTCCTGAC TCCCCCCC GROOCTATITI MICTOCCC TOCTOGACAC TOCACCCACC GIGICATOTT **AGACTICCCCC** 1601 5101 6301 6501 6901 7001 1017 1017 1301 7401 7601 8001 8101 8301 5201 5301 5 5501 5601 5701 5801 5901 6001 8201 8401 8501 8601 1018 200 9 <u>2</u> <u>8</u> 6101 5201 6401 6601 6701 6801 101 7901 8901 9101 9201 8801 900 9301 9401

٠.

12600 13200 13300 13500 12300 12400 12800 12900 13000 13100 13600 10700 10800 10900 11000 11100 11200 11300 1400 11500 11600 11700 11800 11900 12000 12100 12200 12500 12700 13400 13700 13800 13900 14000 14100 14200 14300 14400 14500 10000 ATCCCCCCAG ATCCTCTTCG CGACGGCGC AAGGTACGTC AGAACAAACG GTGCCTCCTA CATTCATTGT GATATCGATG CGACTOCCOG ACTTGACGCC CACAGGAGCG TACGCACCGG GTCGCTGGAA CCCAACGACC ACGTACAAGG CCTTCGATTG GOTGAATTAT CAGAACTTCG CAGTCACGAA CGACGGGGGTG ATTTTCTTCT **OCCAOCIACCO** TOCGATCCAC CCTCCCCCAA CAACAAATGG **ANATICICETTIC GCTCCGAACT** TCAACCTGCA **GCATCGACGA** ATATTCCCTA TTCAGTTCCA ACCCCCTCAC GACCAACCAG CCACCACCGA CACCGACTGG TYGAGTCTGG TCAGGCATCG TCGTCAGCAG AATTTCGACG ACTTCANAGG **GCCACATACC** TGAACTCAGC **GCCGGGAGCT GCAGGACCCT COCFFICTION** OCCUTATATICS TOCTOGGATO AAGAACCAAT **GCCTCGCCGT AMACCITTC** TOCTOCTOCC CTACATCCCT TCCAGGATTC GCGGCTTGGG AATGTCAGAA TCATCOCCA **CCCCCTCCCC** CHOCCCOTCC CCGAGATCAC COCCTTCAAC CACCCACCCG ACCCCCCCC TTGTCGCCCA CCCCGGACAG GTCCACCATC GGTGACATTG CCGCCGATCT **GCCCCCCCAC** CACCACCTTT CCCTGCGCGC TCCCCTCCTC ACCAGTOCCC ATGATCGACG GCGGGGTAAG CTCACCAACA caceroceae ergaaceebe CTGCAAACTA CTCCCGGAGG OCCCAAGACC TACTOCGAGG ACCANTIGAA TATCACCTCG GICTTCCCCA TTGTGCAAGG CCCCCATCGA CTCGATGCTG GCCTAGACTC ACCCACTICCE GTGACTCACC TOCAGCOCCE GCATCGACAG ACACCCTTCG TOCOGTCGTC GCCCGCGGA CGAGAGGATA TCACCCACGC CATTACCGCG TCCATCGCTT GGGGGTGTT COCCGAGCAA CTGTTCAGTG TGAAGACGCT GACGTTTCTC GTACCCATTC ACCCCTTTCG AATCCATCGC GAACCCCCCC GAAACOCTAC CTTCCCGAACG ATCATCGTCG TCGGAGCCCT GCACGTAGTG COCTGATAAA GCCGTCGTGC ATGCTCGAGC CCATCATCAG CONCOCCCC ACCCCTVGC ATCOTGCCA TYCOCAATOC TAYGAGACCG CTGACCOCOC TCCCCAGCCC AAACTOOCAT ACTACCCCG TCACCACCC ACTITACTACC CCGACCAGAA CATCOGGCTC CCCTCTCCC CCTTCCTTCC GTACCGATGC GITARCOGAC ARCODOSTIC TOCCGACACG CITISTCACCC PROCEETING COCTOCOGIT CGACAACCTG CACCTGACCG CTCTGCGGCG AGICTOCTAC GACTACCCCT GCATCTTTCT AGTGCCCCCT COCCTTOCCC COCCAACATT GICGICANAT OCCADOCCOA ACACCOCCAG CICATACCTO GTTCCCCACC GACGTCCCAT OCCCACCCCC TCGAGCACCC CGATGATCTT CTGTTTGACC CCCTTCCACC OCOOCACCOA TOTCOSTCAG CAAGACGTTC AAGAGAAACC CTGANGTAGG TATTICACGCC ACCCGANAGG TATO SCINCE CCANATCACT CONCCCTT TTGGCAAACG ATCTTCTCGA CCCTGTGCTC OCCGACCTGA CCGACGCCGA ACGACGAGTT COUCTOGACC COCCOGIATE TTECTAACTE CAGGIGAGGT AACCGAACCT CAACCCTAGG CAAGGCATTC COTOCINGO ATACAGOOG ACCCCGTACT GITGGCGCGA TCTACTTGCT CACCCTAACG TOCCGAGGTC TCAGGCAGTA GAACATGCTG GCTTATICTICS ACTACTACCG AATCATTCGC TCAACGACAG COCTITICTICC CGTTTGAGGT CTTOCACAAC TCATCOCTAA ATACCAATGC TOCOCCCAN TOMATCOCCC **GCAATICAATT** ACCTCCCATC ACCCACCAAC ACACCGATCC OCTOCCAGTC CCACCCTCCC CACGACTTGT CAGCCTGCCC CAGAACOCCG GCOGCACOCA CCCCCCAT COCACCAAGT AAGACCGGAT CACCETOCCC CAGCGACACG AAGAATAAAA **GCGGGTTGCAC** CTCTCCAACG CTACACCTCG AGCGCAGATC ATCGTCGACT ACAAATGTCA ACGAGGTTTG TOCATOGICA CGTAATOGGC ATCTOCTAGE occoocong TOCOGNICTOS CCACOOCCOC GAACCCCCT ACTOGREGATE OCCCANGGAC AAAGATTCCG ACCCCCACC ACCCCAAGAG CCACTATOCG CAGCCTGCAT GAIGCTCATA **COCCCCATICG** CCCCCTGCG ACATCAACAT CACTCCACGC CICCCACCIT GAGCCGATGGC CCCAGCTCTC TOCCATOOCC ACCANGACG TCCCTCATCT CCANCCACCA CACCOCCCCC TCACCTCCCC ACCOUNTICA CTCCCCATGT COCCCTCCCA TACCCCCCAG ATGACCOOCA **GCTTGCCCGCA** TOCACCOAN CCACCACAT TCCCCCTCCG CCCCACACCG CCGCCCTACA CCACATCGTC TOCCACCTG CCGAATGCCG GCTTGGACCC CCAGGTATTG TCCCTGCTAG CGATAATCGG TTTTCCCCCC 1GTCCTCCCCC GATCGCCATC TATGTCGGTG ACTOCOCTTIC CCCCACAAGG CCTGGGCCCCA CAAGTGGGCC TACCCAACCC GATCTTCTCG **CCCCCATCG** TTTOGACCCC CTCATCGCAT GIOGRAPICOS TROSOCTOSC ACCCCCCATG COCACTICTAG GAAGAGAGCG ACCICCCOTO CCANANGTCC ACCCCCCCCCA OCCUPATION OF TACCCCATTIC COCCCGAMAT COCCTTCTAC TACATCCCCT CCTCCCATAC MATCCMATCT **acretemen** CAAGGTCTAC Terreseree TIGICOCOTT GNACCOCAGA ACACAACOCC COCATOTOTT CCTCACACCT TOCTCAGGGT TOCACGATCC CHOCCOCCAC CCAGCGCCC CCTCCTCCAC GACTCACTCC ACCOCCTICG TCACACCCTC TTTGGGCGG CGGAAGTGAG CCTCAACAAC OCCUATOCCT TCATCGTGGT CATCOCITTIG CAGGCGACCG ATCCCTGTTT ACANGGAGGA **CCTANCITIT** CCCCCTTICAC **GOCCTTCATG** AACCCTTACC TOTACTACTC CCACCCCTAC OCTOCTOCAC ACOCTCTACG COCTACCCAA CCACTITITIC TCGAGCACCA GTCGACTCAG ATATCGCGGC CTOTOGRAFIT **AATGTCCGAC** CTCAGCCCCC CCCTCCTCCT CCCCCCCAAC GTTAGAGGAT COCATCARCT CTCTTCGGCA ACCDGGCGAT CTOCCCACGA CCCAGAGCTA TACCGATCGC GAMACTOCCC **GCACCOCCOC** COCCUCAT CCAACCCCCA GENOCOCCAG TACCACCACC TOCOTOCTICS OCTOTACCTC GACGATTTTC TCGTGCAGGT COCTCTCCAA 1CCCCTCTCG ACTTICGACCC ANGATOGCAC AGCOGGIGCO CACCOTCGAC MOMATCOC TTCCATCCTC CCATACCCAA ACCETCAATG **OCANGOCAGG** CANTCACCTA COTCATOCTC CCACCCAGCA CCTCATCGCG ATCCCCCCTC TGACCAACAT GTCCGAACAG CAGRICCTRCC ATGAAGCCGT COCHEMICEA GTCCCCCCCC CTOCTCACOG **OCCAGATTCA** CCACANGTTC CCTACCCAGG ATCCATACCG CCCCAAGCAG TCTATOGTOC TTACCTOCCA TTGGGCCCCC GAGTATGGCG CCCACACCCGT CTCCTTCCCT TTCCCAATTA TCATCATCGT TOTTGCGATC GTCATCACGT COCTCACCTC CCAACCGTCT ATCACTTOOG TCCAGTGCGT TOCCACTET ACCCCCCTC ACACCCCAGG ACCAGCTGGA ACCITITICACC COCCUCITICA OCCANCANGG **OCCUPACTIC** CCCCTCATGT ACTIGCOCCICA CATOGETCEC CACTCOCTAG ATTCCGTTGG CCATTACCAT OCCOCOCOCC CCCTAACCGT CATGACCATG CACCCCCTCC CCTTCCACTG **GCTANTECTE** ATCATCCCAG CCCCCTCCAG CATCTTCATG CCCAGACGCG CACAACAGGT CCCTAGAGAA ACCOCCCCAT COCTICTICCC TOGATCATCA CCGCAACCTG CCCCCCTCCC TTCGTCCAGT TOTTACCCCT 900000000 **OCTOOCGTCG** GREETERE COCCTATICCG GTCCAGGTAC CCTCACCCA CTTCACCOCTC ATOCACACOG ACGTIGTICGAG OCCATGMAG ACCITCHOCCA GACGACCATC ACACCTGCGA CCCTTACAAC COCCCADCCC COCCATOCCC TATCTTTGTC **CCCCACCCT** ACCCCCCCT TATCCCAAGC CCACTTCCTG TCTCCACCT COCCTATIGIT COCCATCAAC CCACGTTCTC TTCGGCATCA ATGTCGCGGC TOCTITICATE **ACCASTAGGT** 100CCTC00C TACCTACCAG ATCCTCATCA CCCCACACA CATECHEGE **OCCADOCOTIC** OCTOGTOCAC CCANCAGCGG AAAGATTTOC CCCAGCAGCA CCCCCACCAA ATCGACGATG **ACCATOSTO** GATCCCCCTG CACCOCCAAT TOCCCTACAT ACCITCTICGAC TOCCCATCAA CACCGATACC AATTACATCG AGTOCOCCAT GGACCAGGCC MACCAGACCG TCAACCCOOG cococrocic MOCTOCOTT TCCCCCCAT COCCTACACA TTCCTCCCA **accracter** COCCOCCTC ATTOCCAAGO ACOCCCCCCA GAACAAGATC ACCACCTOCC GATCGCCCCA ACCCACAAGG ACTICICGAL TYCCCGGATC CCATOOCAGE CCCACCCOCC COTCGATTOC GATCCTCACC ACAAGTTTCA **OCCUPATION** COTOCITAGE CACCOCTITIC COCCCATGOG GTCAGCGTAC GACTTCTAAC CAAGAATCGT 1301 11701 11801 11901 12001 13101 3201 3301 13801 100 1301 1001 11201 11401 11601 12101 12301 12401 12501 12601 12701 12801 12901 3001 3403 13601 13701 13901 100 1021 10001 0201 1000 1000 10501 10901 10701 10801 080 11501 12201 13501 14401 1051 14601 4701 9901

7/16

14901 TGAT 15001 TGC	GTATCT TOCT. GAGGCA CCAC	AGTATC CTATA GATCAG OCATC	MOCOCO GOOCO	ANTONG CITCTIC XCCGAC GOOCG	CTCCC T	TOCCOCCC ACT	14901 TRATGRATCH TECHNOTATE CHATAGOGG GOCCONTOTO CHITCHOCOCC ATCHONCOC CHICANTICS GOCCOCCOC TON 15100 15001 TONCONOCA CONCANTOR GOLDCOCCOCC GOCCOCCOC GOCCOCCOCC ATCHONCO CCTICANTICS GOCCOCCOCC CONCANCICE TO 15100	CC 15000
15201 COCT	באסכומא נססכ	15201 GOTTCACCA COCTCATOR COTTCCCC GOCGATTC	SCOCK COKEN	ACANCT TOCTO		ACCTIGING THE	15101 OCCUARCOCA CANDITICA CONTROLLA CONTROLLA TACTOCATA GARCOTOTAG TITITACIAGE COCCACTOCA CONTROLLA CANDITICA 15200 15201 ORCIGACIA COCCICATOR COTTOCOCAC GACGANTIC	CA 15200 15239
	<u> </u>	20 30 40	_ e	- 00	20	09	1 1 06 1 08 1 02 1	100

Page 1 of 3

9/16

2000 2700 2800 2900 3000 2300 2500 3100 3200 3300 3400 1000 1200 1400 1600 1700 2100 2200 2400 ACATCGTCAA 2600 3500 3600 3700 3800 000 1100 GACGCACCGA CATCACCGAC CTCACCCTGG 1300 TOCCCCCC 1500 GACCTTGACC TOGGGACACC CCAGGGCCGG CTAGTCGCCC GCCTGAAGGG 1800 GAAAGCCTTC 1900 OCTACCCCCA 1100 AGAGACCTICC CAGGAGGCGA 3900 GAGGIGITGA CCACCCCGGA ACGGCTGCGG TCTCTGGAAC 200 GAACTIGGGCG GCACGCTGTG 300 OCCOGGCTG 900 TOCANGECET OCCARGOGET GOCCATECTIC GCCGACAACA TCGAACCGAT **OCCANATICC** AGTECITIETE CACCEGOTEG GCAAGAGEGG CAGGATATTE AATCECGAAC ACCTOCAMAC ACTIGITACIANG AACGTTICCOC OCCTACTIGAC COGACTOGCC CONCOCCATG TOXCTISTIGGE GEACCETICTE CGAAAGCGAC ACGCCGAACA GETTICGACTE COCCETITIOS GIETECECEAA GEAGTECECAE GTGTCGACEA GACCATCCCG AGITICATCGC COCGAAAACC GATCTCGTTG CCGCGAACAT COCCACACCC CTGATCGGCG AAGGCGCACA TCAAAGTGAT TCGCGCCCTT TTTCGGCCCA AGTACCTOGA CAACGACGTC ACCCCATCAA COGGCAAGCG CCGCCCCCCC OCTOSTCAAA CAGGCCTACG CCGACATCCT CGCCGGGGCG TCCCTGGGCG CCCTOCACCA CTACAACCCT GTCGAAATTC TTCCCCAAAC CCCGCAACGC CGATTOTICOS CAAGOCCCAG TOSTCOCCOC TOSTOGACGA GOCGACOTTC TOTTOCOCCO OCAGATTICAC GACGCCCCCG AAGCCGAAAC CATCCCCTG GACTOTTICGA COCCATACCO TTCCCAACAC CCCAAGTCCC CCCCATGATA GTCACCACGG CCCCAAAGGC GATGCCCCCA TCGAACACGA GCGCTGCGAT CCCCCAGCCC GACOGGIOGG TOCOOCTGCA CTOCCAATTIC GCCTGCCTGA OCTACGGCGC GCCCCCTTCCG GCGCTCAGGC STYCOCCACAC CGTCTAGCGC TCACTCCCC ACCTIATCAT TCCAACATGT TACCCAATAG CCGGGAGGAG CAATACGACG TCGCCAACGA CCGCCGCTGC TCCATCGCCG TCCCATCGAG CTGGAAGCCT AAACCCGACCC CCAATTACGC CTCAAATTOC GOCCCCACGC GCCCGCTGAC CGCATCACAA AGATATGCCG GOCGCGACCA ACCCCGACGA **GOCGCCCGCCA** CCACCCCACC TCGACCACGG **COCCOCCT** CCTCCCACCC CACCTCCACA TCACCGGCGA COCACCAGCT GCCCCCCTAC GOCGCCAGCG GTGGCAACAT TACTCCACCA CGACAACCAG GAAGGOCTAC TGACCGCGC CCAGGTGAAG ATCAGCACCG COCOGCCACC CCAACTOGTC GACCAACCAG CTCCCCCCAA CAGCTCCCCC TCAACCACAA GACGCTAGCC GTCCCAGCAC COCAACCACG ACGGGCTGCT CCTCCAAACC GCAAGCGGGA **GOCAGGCTGC** GCCCCATGCC TACATCCCCC CCACCCTOCC COCTOCCACT AACCAACTOG ACACCCAAGC CAGCGAGGAA CCATCOCCGA COCCOCCGAT CTCGGACCTC GACCTICOCCG CAAACCGCTC AAATATCGTC CCATCACCCT GANGCOSTOC TAGCCAAACT GOCCGCCCCC CTCGATCGTG GTCACCACCA CCCTGACCGA CTACTCCCCC GOCCCAGCGG ATCATGCTGT AACACCCACG GCCACACCGA ATGGCTACCA TCACCGGGCT AGCAGCGCAC CCGGCTGGCC GAGCTGCTGC TODOCCACOG AGGACCOGGC ACCCGCGACA accascing CTCTCACCCT **OCCOOCCETC** GATCCACCOO GACOCOCACO TOCTCATCGA COCCOCCATIC CAGACCTOCG CCAACCCCTC TCACACACCC CATTCCCCCC OCCCOCCAGA TOGTCCCACG CACCOCCTOS ACCAGCACCO CAGAAGACCG CCCCCCCCC GAAACTGAA AAGCGTGCGC CGCCACCTGC CONCIONACE COCONCIDOS TOCTOCACGE CCTGGCCGCC GCACTGGTAC COCMARCOCC ACACCCCCAG CCAAGCCCAA ATCACCAGCC ACGCCCACCA CTCCACCCCC CCTCCATCCT CONCINENT OCCTOCCAC CAGGACAAGC CCAACOCOCC TACCTCCCC CTCCCAATCT TOCOCOCOST GAATTCACTT ACCTAACACC ACTTCTAGCA GCTGTCGGCG CGACTTCTTG OCTOGATOCC GAOCTOGACC CCTTGGACGA CTCCCCGAAA AACCCTCGAC CACCCCCAAA TOTOCOGOCA OCCANATOS GTOCCOGTOS CCOGCCCCAA COGTCCCCAA CCCCAACCCC GGACAGCGCC PROCEEDETA COOCCEDETE GACCIGICACE GOCCOCOCC CCGCCCCAA OCTAGACTICE GACGTAGCCG GTOCTCGGTG CACGAGCATG CACAGCCAAG CCCACCACGT COCANGATOR COCCOSTOCT CCCTOCCAC CGTCGCCGGC OCTUALISAMA CYCGAGCACAA GAAGGCACGA CAGCGCCCCC COTTCACCAT CACCGGCCGC ACCOCCAGGT COTCTACGTG OCCIPOCOCT CGAACOCOCC CCCAACCCCC TOTATION COCCOATOR CTCACTCACC COCAMIDOS CCOCCATOTO COCTACCOG COSTCOGOCA CACGITICATO CTOCOGGETO OCCANOCOGY TACOCATCAC CANOCOGGAC GCCGCCCTAC CCACCOCCCA ACOCCAGGG GIGICCAAAC CCOCCAGGCC GCCGAAGCCC CANCECCANCE TCANCCEACAC CCCAACCC GCCACCTIT CCACCCCCGA TOCOOCCOCC ATCCACCOCG CLACACAACG GITCTTCCCGT CCATOCCCCA TOTGATICCGC GACCACOCCA CACCCCTOCC OCTICTATICAC ACCAAACOCC CTGTGCGTGC TGCTGTCTAC CATOCCCGAC GCCGTTCACT CECCAMENTA ACCAMACTIVE MOCCCECCA ACAGGATICAG GAACGGCTCC GICCTCGACG TCCTCGCTGA CCACAACATC GTCATCGTCT CCTTACCCCA TACGACCTCA AATGCGCCCCC GTOCAGTICOG ACOCOGTOGT COCCOCCGAC GCGACGATGC CCCCCCGACTG GCCCGGCATC GACTATCCCO ACCCCCCTG ACCOSCITAGE CGAACGITAC ACCANACGT GCAGTOCTGG accordance econocinea CITICCTCAAC GTCGCGAATG GGACGCTGGA CCTGCACACG COTOCOGGITT ACCCCCCCC ACCCAGGGCT CCACAGITIG ACCOCCACCO COCOCOCTGA TCCCTCCGG GGAACTGGGC OCTOTICACOC ACCOCCTAC TOTTOCCCGA CATCACCGCC ACCCGACGAG ANGCOCCTOG CCAGTOGAAC GACGCCGGGG AGCCCGTGCT GGACGCCCCC TACCOCACCG CCTGTATCAC ATCGTGGCCG AGCGGCTGGC GACCTCCACA GAATTOOCCA TGAOCCGOCA CCTCCACCAC OCCOCCCAAC GAGTCTCGCC OCCAACCCC ACTACCCACC COCTTTTGCG COTOGTAGCT GTAGCCGCAC CCCTCCCCCC ACCTICIANCIA COCCIATIONT CTACCOCACC CGAATCCCCT CAGACGCTGG GCAGCCGACG ACCGCGGGA GANANTICCEA ATCOCCTCC ACCCTGCTAC TACACTACAC GFTCCCCCCC COCACOTIGGG AGCAGCGGCC accescent CCCTICTICCTIC GTATCTGCGG TCATGGACTG OCTACACCCC OCATIGICACO OCTAAGTICOC TACCTICACCC CITICICCAGO ICITICIATOC GICTIOGAATO CITIOGIOCOC COCCOGTOCAT GTCATCGACA 000000000 CCTGCGACCC CCACAACCGA ACCTIGTICAGA TOCCTANGC **OCATATANGG** GOCCGGCAGC CCCTCTACGG CCCCCACCC CCATCATTTG CACCCCAACG TCGCGGCCGC CTCATCCGCA GACCGCCCCCC **OCCITCACOCO GICACGICAC** CCCACATCCC COCCUNGCC ACCOCTAGCC ACCTGCCCCC **OCCCAGCGGG** COCCATCTTC ACCAMACCCG CCAACCCTTGG ACCCCAGGAC TACCAGCAGA COCTACCTCC ACCITCACCCG COCACTACOC TOCOCCOCCC COCCOOCCIG TCTACCATCA **GCTGCCCCAT accreaecete** OCCUPACETICC **GCGACCCCGCA** CCACACCCCC ACCCTATICAC TCATGTCATT GICCOTOCCC **CCAACCACCT** GAACTOGAAA **OCCCADCTGT** COCTOCACCT TTATTGGCTT **NACTACGCGC** COCCOCACTAC CGATCCTACC TOCCOCNOCC COCCUCATOS AAGTTCATCO CITICICOCOC 3401 1201 1301 1401 1501 1601 1701 1801 1901 2001 2101 2201 2301 2401 2501 2601 2701 2801 2901 3001 3101 3201 3301 3501 3601 3701 3801 3901 1001 101 1021 200 5 501 50 5 801 1101 901 00

# Figure 3 continued, page 2 of 3

5700 5600 5800 5900 0009 6100 6200 6300 6400 6500 0099 6700 6800 0069 7000 7100 7200 7300 7400 7500 7600 7700 7800 8600 8000 8100 8200 8300 8400 8500 8700 GCGTTATTICA GCCGCTCACG 7900 8800 8900 9000 9200 9100 9300 Nenesces internatel Managana Managanan 9400 **ATTCCCTATG** TOCCATOCCT CTACGCGCGC GCACGTATGG ACCCTCACTG CACCCCACTG GCGAACCCGT CATGGGTGTC ACACCATICCT CCACTCCCAC CCGGCGCAC CCTACGACTC CAAAGCGTTG TTCCCCCCCC ACCTICAATGC CCACAGTACC GACGACACCA TTGTCTQCTC GCAGACQCCT CAACCCCCAG CGACCTATCG AGCAGGACGA OCCOCCATOC OCCACAAGCG ATTCCGCGCG TOCCOCCTUC GTTCCCCCCC **AAATTICATOC** CTCACAATTA OCCANAGE GNACEGIBEN COCACCATE ATTOGGGG ACAAATCGCC GOGGTGCGCG CTATGACCGA ATTCGACGAC ATCAAAAACC ACATOOCTTT CCGAGTCGCC CCTTICACCTIC TTCCCCTCAC **GTCTCATACA** CAGCGCCCCC TOCCCCTT CCCCCACCGT GROCICCCG ACCCCTGAT COCCACCOC OCTOCOGGG G GTCATCCCTC TOGGECCEAN TECHEGOCCG GTACTCCGAA CTCCAGGGTG **GTTCCACTCC** OCCCTCAAAT GGAGACTACC OCCTACKTAG NEGREGISM INSTRUMENT MERRICAC ANCCAGANT CCACCCCCC CATTIGITIES OCTOCOCCTA ACCIOCICACIC CCACCAATCA COTTCOTOCA OCCUARGGAA AATCTGTGTT CGACAAGGCG CTCCGCGGCG GCCATCCGCG CCCGAAAGAT TTCCACCCC ATTACCCCC GTCCCNAACC ACCOCCCCCA GTCCCATTCG CTGCAACAGC COCTOCCCGA ATOCCOCCOC GTCOCACCGA CTTTGGAGCG CGTCATCCTC CCAATCCCCG COCCITACCITO CCTAGACTAA GCTGCGCTTG GCCTCCCAGT CTTTTCGGG CTCCATCGCC ACCOCOCCOG TGACCTGAGG COCAMAGICC COCCCANGOC GACCOGTGAT GACCTTTTIGT GGAGGAGCCT CACGGCTTGT TOGGIOGOTT TOCGACCGAT TOTOTOCCAG TOCCOCCAC CTCCTCTCCC GGATTCCGTT CTCCATCGAG AGACCOCCOC GITCGTCAAC GGCTCCGCCA CTCCCCTCAC AACCTCCTTG CCCCTCACAG CCATTCGTNN ATCTTTACGC ACCCCCANAC TTCCANTGCC CATOCCCAOC **GENERACIO**CO ACCCITICCAG ACCCMACAGE CCAMATOCIAE **GCAACCTCGA** CACCCCGACG CCCCAACCGG ATGCCCGATA TGTTCGAGGC TCACAAGGCC CCACCTTCCC CTOCCOTTC TOSTTCATOS ACTIGGGGGGA CTGCAGACCG ACCONCICCO **GCACTCCAGG** AACACTTATC CTCCCCCCC ACACCGAGAA GCGAGACGCC ACCCTCATTA GOOCOGITIC ACCOCCITECG TCGGCACAGC CCCACTOOCC TTCCCCCCCC ACCCCCCCCA ACTOGATTICT CAGCGCTGCC CCCONCGATC CAACCCCCCC GTGCCGCGG AAGCCCCCC CGTTGATTCG COCTIONAGE TOCCOGOCOGA CCTCACCTGT OCCUPACING TICHONAGGE GITOCGITICI CATCCOOCOG COCCCCATICAT COCCCCATGA CADGACCOGG AACTOGACGC CONCINCACO TOCCOCCAC ACTACCCATC ACCCOGGICAC TCACCCATTT GCCCAGGATG CGGTACCTAC CATCCOOCCC CACCCCCTCC GAMACCCCC ACCACGCCGA CCTGCCCCAAC GTGCCCGTTC CTAATCGCCG OCACCAGCTC TATCCCOCCA ACGACCCOCT CACGGTGCCG GCGCGCATGC CACCAGTCCA COCCOCTCCC CCCCCCCCAA **GCATGCATCT** GTTTOCCCC COCAGTECAC CTEGTODOCG CCCCATAGCG **ACTIONOSICT** OCT CCT CCCC TTCAGTTCCG CCACTCGTTT OCTCAACAAT CTOCCCCCT ACCCACCCCC COCCOCTOCA COTCACTTCC GTTCCCCCCT CTCACCCCTC COTTTCCCAC **GCAAGCTGCT** ACAGCCATT ACCTTCCCCC ACATOGACAC CCACTTOCTC TCCCCTOCOG CCAGCTCGC GACAACTCAT COCTGTAGGT GAMACOCTC MCCCCTCAC AGCCGCACGC GAMACCACT ACCCACCCAT **GTCACCCATC** ATCACCCCCC OCTOCOCCA CCCCCCCCAA CCTACATCAT COTCTACCTC COCTCAACCC COCCTOCAAC CTOCHETECT **OCCUPICATE AC** CCACCACCAC COCCACCC CACCETTTOC GAGACOCCCC CCATCCCCCA **OCAMATGTTC** CTACTCCATT 2000000000 CCTAACCGC CACCAACTOC GCCCCCCCTC CTGATCCCCC TTTGGCGGGG GAACTACCGT CACCATCAAC GAAGTCTCCC **GOTCCATGGT** DOTOCOCINC COCCOCCANC **プラウラウラウラン** TCCCCCTT TCGFFFTGAGC TCTCTCCCA CCGTTCTTGA CATGICCTOG CGATTICTTAT OCTOOCCEAN TCANCGATAN TAGTGATTCC TCCCCACCAG TTATOCCTOC ACCOCTGAGE CTGACCTTTT CATGCACCG **GCACTATCCC** COCCOGNOCT GTCCCTCCAC CACACCTAAC GROCCOMCT COCTIGECOCG **CCCCCCTCCT** GACCCCCCCGA CAMAGTOCTG ACCEPTORING ANAGACACA ACCEATGCET CGATCAACCG GCCCTGACC AACGACGAAC CATCGTGACT COCTTOCCTT CCACGCGTCG CATCGTGACC AGGCCCAGCC COGACACCGC OCTUCTIGATE GOSTACTICES AGACCTICATE GACCCCCTCC TCCCTAAGCC CCAAACACCC ACTOSTICAG GACCCCCAGG GACCOTTCCC GACCICCCACC **GCGAGTTCCA** ATCCTTTTCC CTACCCGGTG **AACCTAAAAG** GCCCCCGAAGC TACACCTTTC CTOCCITCITC CGACAGIGIC CAACTCGTTG Accooccoc TCCAATCCTG TTGCGGGGCA ACCCCCCTTT ACCTCCCACA Greencence COCCOMOCC OCACAOCOTT CGAACTTGTC GICTIGICCTG GACAGAGTTG Personal conservation NACESSEE NAMES OF THE PARTY OF TCCCCGAACA CACCGATTTC TITIGTCGTCG CTCCCCCTTG CGTTGTCGAT CACATIGNOCT COCCOCCAC TCACCGAATC STCACCAGCT CONTRODUCA CCATCATTAG ACCICITACC **OCCOCATIGAC** CCCCCCCTAT ACCETCACCC CCCGTCCACG CGAAGCACTC GAATCCTTCG CCGNAACCCC CCCCCCTCAA GOCTITATIC CGACCTGCAA ACGCTGTGGT COGNIDECCE GGACTOOCTE ACCCGGCGCT COCTITICOCC GACCICCION COCCOCCOC **GCCCGGCCGAC OCCODOCAGT** TITICAGTACC CACTTCCACC CTCGAACCCG **accecentar** ACCACCTCCT COCCOCCOCC CCCCCGGAAAC CTTGTGCCCCC ATGGGCCTGG TGGACACCCA **GOTOCANA**CG CACCCCAAGC TTGGCCCACC ACATTICTICOC CTCCCATCAG **CCCTTCCCCC** TACCCOCCAA CCTCATCGTC ACCIDENCE CCTGTTGCCC ACCORTCOCAT CAGTECCACA CCGAATCGCC TACTAGGCAC COTCCGCGAA CCACCCACTA MGATCGCCG CCGTTTCAAG CCACCAATGC TCCCCCAAAG CCCATCATTT CACATOCOTC CACCOGCGGT AGTCCTCGAT CCACCOCCAC **OCCUPIENCY** CCCCAACTGT CTCCACCCCA TOCCOCTOCC ATCCCCTCAC GTCCCCCAAT TCTCACCACC CCCCTCCAA GACCCCCCCCA OCCOCHOCHO CACCGTCCCC ACCIOCCACCA TCCCGTTCCA CAACCCCCTT TTCCCCACC Nerstand Nerstand TOCCGACCCT CACATOCOCA TAACCOOCAC COCCACCCC CONCOUNT OCCIOCITY INDEPENDENT PROPERTIES NESSESSES TCCCCCCTT CCCCCCATCC COCCIDENTIC CACACCCANA TCCAGTOGTO GTCCCCCAT **GCTTCATCGA** ACCCGAAATC MCCAGCAM **STCGGTATOG** AATCCAACCG GTGAGAGGG TOGITICATICA ACCTGCCCTA **GCCCANACGT** OCCIONACIOC **OCACTTACCA** CCAGTGACCC TOTOCATOCO OCCANGACCA ATOCCOCCOCT acceptance CCTCACCAC AGGATCGGGC ATCCCCTCTC **OCATCTCOOC** ATGTGCTGGT TOCCCCCCA CCACCCAGCG CACCACCAC CTOCTAGGTG ACCAGGCCAC **CONCINCIA** GACACCCCOG TCAMACCOCT ACCAACTOCO TACAGICGIT COCCOCCAC TOTOCOLOGA **GCGCCCGAGA OCOCCTTCGT** ATTCCCCACC CTTCGTTGGC GTCAGCCCCC CARCETTCCA CCACCACACT CTTOCCCACT **GCGCATTICITY** COCCUTOCOCC OCCOCCATOG COCTOCCTAC CCCTTOCCCA GTCCTGGGCG ACCOUNTE TYCCTOCAAT ACCATANACA ANGANGOCGE COCAATAGCG TOTACCOCCA TOCCATOGIG CTOCCCOCCO TOCTCACCCG **COCCITCOOCC** COCTTATACC TCCCACCCCA CCACCAGGG CCCCAGAACC CACCCCAT **OCCCACANCA** TACAGGTCCA **OCMATIGITICE** GACCACCCAT GANGTITIATIC CTCANACCTG CCTCACCGA 200000000 ATCCARGITITO TITCHIGGE ACMATTICAN CCACATTCCC CCCCAACCC COCAATCGCT TCTCTTTTACC COCCACCCC GTGCCCTACA CCCTCCCAC GTCCTCCCCC CACTATTICS GACGAACCCG **OCCCADGAGG** ACCCCCCCCAG **OCCUPACIOC** TOCCATOCOT TOCACTOGIC ACCEDECAGE GATAGGGCCA ATCCAOCCTG CACCOGGITTC 5201 5301 5401 180 1901 5001 5601 5701 501 5801 6001 100 7601 590 6101 6301 5 6501 6701 5901 100 1201 1701 8101 8201 6201 5601 5801 5 7501 7801 7901 8001 8301 8401 3501 8701 **B601** 9801 1901 1000 9101 9201 3301 5

10/16

9800 10000 10100 10100 10200 10300 10400 10700 10700 11100 11100 11100 11100 11100 11100 11100 11100 11100 11100 11100 11100 11100 11100 11100 11100 1100 1100	
ECTC  THE	100
THOOCOCTC THOOCOCTC THOOCOCTCOAN THE COCCOAN THE COCCO	-
CCA	90
ALTUMAGACIA CCACCOTOCO GTROACOCTO OBATTITUTI GEOCOTOCO TOGATITUCOST TTACCOTOGA TOGACOCTO ATTOGATO TTACCOTOCO ACCOCOCOO COCOCOTOCO COCOCOTOCO ACCOCOCOO COCOCOCOO COCOCOTOCO COCOCOCOO COCOCOCOCO COCOCOCOO COCOCOCO	-
	80
A.T.C.AGGGGGA COCCOCTOTORA TITGCCTTTCG COCCOCTOTOG GCCCGAGCA ACCCGAGCA TITCCAGGT CCCCGCACCA CCCCGAGCA TCCCCGATT CCCCCGAGCA CCCCCGAGCA CCCCCGAGCA CCCCCGAGCA CCCCCGAGCA CCCCCGAGCA CCCCCGAGCA CCCCCGAGCA CCCCCGACCA CCCCCGACCA CCCCCGACCA CCCCCGACCA CCCCCGACCA CCCCCGACCA CCCCCGACCA CCCCCGACCA CCCCCGACCA CCCCCGACCA CCCCCCACCA CCCCCCACCA CCCCCCACCA CCCCCC	-
ATTEC COOKE CO	70
CCGGCATTCG TCATCCATGG CCGCCCATAGG CCCCCATAGG CCCCCTCACG CCCCCATAGG CCCCCCCCCC	-
THOSE COLOR OF THE	9
TTOTTTCTCC GCCCCATAG GCCCCATAG GCCCGATAG GCCGATCATC GCCGCCCGC GCCATTCAACAG CTTCAACAG GCGATCAACC ACCTCAATC ACCTCAATC ACCTCAATC ACCTCAATC ACCTCAATC ACCTCAATC ACCTCAATC CCAATCAACC CCAATCAACC CCACCTCATC ACCTCAACC CCACCTCATC CCACCTCATC CCACCTCATC CCACCTCATC CCACCTCATC CCACCTCATC CCACCTCATC	-
MANGO COCHA	20
TOCOCCAAAO COCCOACAA COCCO	-
CCAA ACACA ACACA ACACA ACCA ACCA ACCA	<b>\$</b>
	-
COCCA COCCA	30
CCGARGACGACT CCGARATATOS CCGARATATOS CCGARACATC CCGARACATC CCCGARACATC CCCGARACATC CCCGARACATC CCCGARACATC CCCGARACATC CCCGARACATC CCCACACACAC ACCCACACAC CCCACACACAC	_
POCOC COCCA AND AND COCCA AND COCCA AND COCCA AND COCCA AND COCCA AND COCCA AND COCCA	20
ACTUTICACO GOGCOTTICOA TITICACOTOCO GOGCOCOANTO TICCAACOCO CTCANCOCOC CTCANCOCOC GOACCACOCO GOCOMOCO GOCO G	_
TOTAL  THE COMPANY	10
CCTCCTTOTC TOTCGATTO GOCGATCATO GOCGATCATO GOCGATCACO GOCGCTCCO GOCGCTCCO ANCTGCTAC ACTGGCCO GOCGCTCCO GOCGCTCCO GOCGCTCCO GOCCTGCC ACTGGCCO GOCCTGGCC CATTOTCAC GOCGCTGGC CATTOTCAC GOCGCCCO GOCGCCCCO GOCGCCCCO GOCGCCCCO GOCGCCCCO GOCCCCCO GOCGCCCCO GOCGCCCCO GOCGCCCCO GOCGCCCCO GOCGCCCCO GOCGCCCCO GOCCCCCCO GOCCCCCCO GOCCCCCCO GOCCCCCCCO GOCCCCCCCO GOCCCCCCCO GOCCCCCCCC	_
9701 9801 10001 10001 10101 10101 10501 10601 11001 11101 11101 11101 11101 11101 11101 11101 11101 11101 11201 11	

							_	_					
		<i>WIIIIII.</i> BCG∆ = ~8.8kb	ORF	Homologue Accession #		246257	U01072	X79562				129506	A08331
	₩ —	+ 9	•	P value		1.48-14	3.08-13	2.36-43				3.68-16	4.06-14
	h Nh St	12 13 14 15 1	1F 1G 1H	Homologies to Predicted Encoded Protein		M.leprae aceA	BCG uraA	M. IUDerculosis esat6				B. subtilis subtilisin Serine proleages	Company America
	N		Ħ	Encoded Protein	1111avnDa)	36	76	200	34	48	50	46	
	xb St (miniminiminimini)	- 60 - 80 - 1	Đ	Possible Ribosome Binding Sites	AGGA (10)	GGA (4)	GAGG (5)	9000	GGA (5)	AGGA (9)	GGA (11)	GAA (5)	
	4X	9 1	5	Start - Stop (base pairs)	889 - 2433	3130 - 4203 3139 - 4203	5075 - 6046	6954 - 8612	10619 - 9663	13328 - 11946	14823 - 13438 14643 - 13438 14541 - 13438	16190 - 14820	
( <del>Q</del> )	Hp Hp	e 1	8	ORF Size (base pairs)	1542	1071	696	1657	954	1380	1386	1368	
Region 1 (16.9 kb)	- H	7~ 1	TA T	ORF M. tuberculosis Codon Usage	yes	yes	yes	yes	yes	yes	yes	yes	
Rec	፫	<del>-</del> 0		ORF	14	18	5	9	<u>"</u>	=	16	Ξ	

Figure 4

Region 2 (15.3 kb)

	<i>'''''''''</i> BCG∆ = ~10.8 kb QRF	Homologue Accession #				P24194	U00015 A00975 U03393		X73226 X17445	A30545	X65104 U04851 Z22594	CCLECY			
<u> </u>	BC	P value				9.9e-47 <1e-5	1.5e-7 -4.e-5		9.9e-146	6.7e-141	3.1e-11 1.4e-08 4.4e-11	6.36.03			
Bg Kp Xb	2G 2H 2I	Homologies to Predicted Encoded Protein				E.coli iciA IysR family	Mieprae cosmid B1620 ORF Cutinases		S.typhimurium RNDPR proUVWX	M.tuberculosis mp164	E. coli gabP permease S. lyphimurium asp permease T. harzianum indal gene				
P	2F.	Encoded Protein (max. ~kDa)	25	16	34	34	22	19	37	24	51	31	35	25	21
Bm Sp Bm	20 2J	Possible Ribosomal Binding Sites	AGGGAG (7)	AGAA (4)	none	AG? (8) GGA (8)	попе	none	AGGA (11)	AAGA (6)	AG (10)	GGAAGA (6)	GAG (10) GGAA (8) GGA (9)	none	AG (10)
H3 Bm P	5L 2C 2k ♣ 2C	Start - Stop (base pairs)		2862 - 3298	3003 - 3590	5187 - 6134 5376 - 6134	6561 - 7217	8036 - 8560	9941 - 10909	11118 - 11783	11965-13407	14221-13376	8259 - 7211 7939 - 7211 7931 - 7211	4992 - 4327	5117 - 4521
	2A 2B2	ORF Size (base pairs)	558	437	588	948	657	522	996	999	1443	846	1050	999	297
S B O	_	M. tuberculosis Codon Usage	yes	yes	yes	yes	yes	yes		yes	yes	yes	yes	yes	yes
_	_	ORF	8	<b>8</b>	797	သွ	2D	7	2F	22	2Н	₹	23	ž	ಸ

Figure 5

(q
2.4 k
3 (1)
on ;
Red

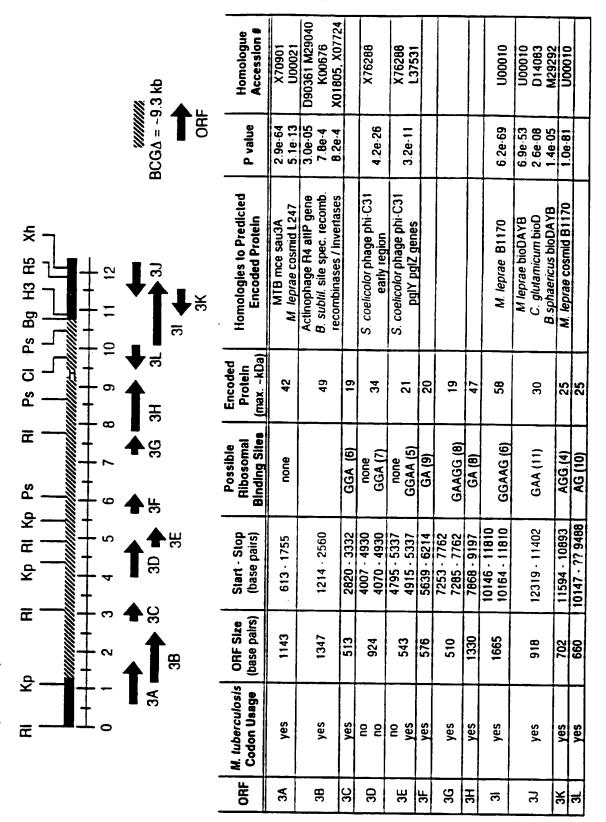


Figure 6

-----CTGGTCGACGATTGGCACAT|GCAGCCGTGGCTGCCGCGG------GTGTCTTCATCGGCTTCCAC|CCAGCCGCCGGATCCAGCA--BCGA1b Junction ---- -16.9 kb - -8.8 kb Region BCGA1a Junction

----CAACTCCACGGCGACCACCC | GCGCCCCCGCTCGCACTAGA--------------GCCCACCCGGTCGAGCACCC | CGATGATCTTCTGTTTGACC---BCGA2b Junction ----- -10.8 kb -------- -15.3 kb ---Region 2 BCGA2a Junction

----CACCTCGACCACGGCCAACC|GTGGACCTGTGAGATACACT-----TCAGCAGTCCACGGCCAACC|CGGACCAACACCTTCCACC---BCGA3b Junction --- -9,3 kb --------- -12.4 kb -----**BCG** \( \mathcal{A} \) 3 BCGA3a Junction

Region

Figure 7

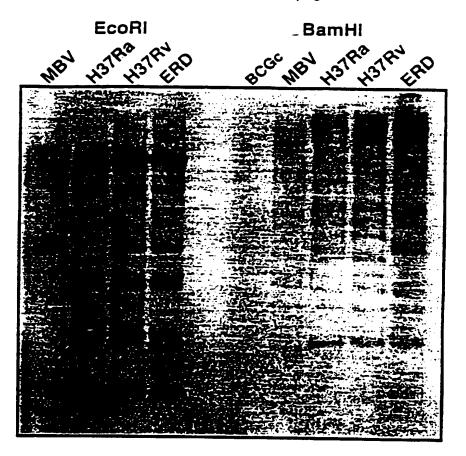


Figure 8

International application No. PCT/US96/01938

A CI	A SCIENCY ATION OF CHIP IDOM A COMPA					
A. CL.	ASSIFICATION OF SUBJECT MATTER :Please See Extra Sheet.					
US CL	:Please See Extra Sheet.					
According	to International Patent Classification (IPC) or to bo	th national classification and IPC				
	LDS SEARCHED					
	documentation searched (classification system follow					
U. <b>S</b> . :	435/6, 7.1, 91.1, 91.2, 240.2, 252.3; 530/300, 3	50, 387.1, 388.1; 536/22.1, 23.1, 24.5, 1	24.32, 24.33			
Documenta	tion searched other than minimum documentation to	the extent that such documents are included	d in the fields searched			
Classas i						
Discussion C	data base consulted during the international search (	name of data base and, where practicable	e, search terms used)			
Piesse 5	ee Extra Sheet.					
			,			
C. DOC	UMENTS CONSIDERED TO BE RELEVANT					
Category*	Citation of document, with indication, when					
	Citation of document, with indication, where	appropriate, of the relevant passages	Relevant to claim No.			
×	1-10, 16, 17,					
	24 25					
Y						
	18-23					
,						
<b>·</b>	1-7					
X JP, 1-247094 (AJINOMOTO ET AL) 02 October 1989, see 1-7 entire document.						
<b>(</b>						
`	Infection and Immunity, Volume 5 1991, C. Parra et al, "Isolat	9, No. 10, issued October	1-17			
ļ	molecular cloning of a specific m	ion, characterization and				
ľ	antigen gene: identification of a s	pecies specific agrees"				
	pages 3411-3417, see entire doc	ument				
		different.				
X Furthe	er documents are listed in the continuation of Box (	See patent family annex.				
	ial categories of cited documents:	"T" later document published efter the inter	mational filing date or priorny			
docu to be	iment defining the general state of the art which is not considered to f particular relevance	date and not in conflict with the applica principle or theory underlying the inve	tion but cited to understand the intion			
· carli	er document published on or after the international filing date	"X" document of particular relevance; the	claimed invention cannot be			
docu cited	ment which may throw doubte on priority claim(s) or which is to establish the publication date of another citation or other	considered novel or cannot be consider when the document is taken alone	od to involve an inventive step			
spec	claimed invention cannot be					
special reason (as specified)  "Y"  document of particular relevance; the claimed invent considered to involve an inventive step when the combined with one or more other such documents, such being obvious to a person skilled in the art						
docu	ment published prior to the international filing date but later than riority date claimed	*&* document member of the same patent f				
<u>-</u>	ctual completion of the international search	Date of mailing of the international sear	rch report			
17 APRIL :		29 MAY 1996	ten report			
		20 IIAI 1336				
me and ma	illing address of the ISA/US r of Patents and Trademarks	Authorized officer	Frut 10			
Box PCT	D.C. 20231	JEFFREY FREDMAN				
csimile No.		Telephone No. (703) 308-0196				
m PCT/ISA	√210 (second sheet)(July 1992)±	(103) 300-0170				

International application No. PCT/US96/01938

C (Continua	tion). DOCUMENTS CONSIDERED TO BE RELEVANT				
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No			
Y	Abstracts of the 1994 IDSA Annual Meeting, Clin. Infect. Dis., Volume 19, issued October 1994, R. Frothingham et al, "Sequence based strain differentiation in the Mycobacterium tuberculosis complex, including rapid identification of M. bovis BCG", page 565, see abstract 10.	1-25			
x	R. GHERNA et al, "AMERICAN TYPE CULTURE COLLECTION: CATALOGUE OF BACTERIA AND PHAGES", Eighteenth edition, published 1992, pages 202 and 211, see entire document.	11-15			
x	Infection and Immunity, Volume 62, No. 4, issued April 1994, L.	1-7, 16-25			
	Pascopella et al, "Use of in vivo complementation in	26			
Y	Mycobacterium tuberculosis to identify a genomic fragment associated with virulence", pages 1313-1319, see entire document.	20			
Y	Science, Volume 261, issued 10 September 1993, S. Arruda et al, "Cloning of an M. Tuberculosis DNA fragment associated with entry and survival inside cells", pages 1454-1457, see entire document.				
x	US,A,5,171,839 (PATARROYO) 15 December 1992, columns 5-	1-10			
Υ	10.	16-23			
Y	Nature, Volume 256, issued 07 August 1975, C. Kohler et al, "Continuous cultures of fused cells secreting antibody of predefined specificity", pages 495-497, see entire document.	10			
Y	US, A, 4,683,202 (MULLIS) 28 July 1987, see entire document.	16-22, 24, 25			
Y	Genomics, Volume 4, issued 1989, D. Wu et al, "The ligation amplification reaction (LAR) amplification of specific DNA sequences using sequential rounds of template directed ligation", pages 560-569, see figure 2.	16-22, 24, 25			
Y	US,A, 4,410,660 (STRAUS) 18 October 1983, columns 14 and 15.	23			
Y	Gene, Volume 131, issued 1993, A. Kinger et al, "Identification and cloning of genes differentially expressed in the virulent strain of mycobacterium tuberculosis", pages 113-117, see page 114, column 2.	1-26			
X,P	WO,A2,95/17511 (JACOBS ET AL) 29 June 1995, see entire	1-26			

International application No.
PCT/US96/01938

		PC1/US96/019.	00							
C (Continue	(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT									
Category*	Citation of document, with indication, where appropriate, of the relevan	nt passages	Relevant to claim No							
X,E	J. Bacteriol., Volume 178, No. 5, issued March 1996, (Mahairas et al, "Molecular analysis of genetic difference mycobacterium bovis BCG and virulent M. bovis", page 1282, see entire document.	es between	1-26							
Y, P	Microbiology, Volume 141, issued 1995, J. Rodriguez e "Species-specific identification of mycobacterium bovis l pages 2131-2138, see entire document.		1-7, 16-22, 24, 25							
X  Y	Hybridoma, Volume 13, No. 1, issued 1994, A. Arya el "Production and characterization of new murine monocle antibodies reactive to mycobacterium tuberculosis", page see page 27, table 1.	onal	8-10  16-23							
			Ü							
			5							
		į								

International application No. PCT/US96/01938

A. CLASSIFICATION OF SUBJECT MATTER: IPC (6):

C12Q 1/68; G01N 33/53; C12P 19/34; C12N 5/10, 1/21; C07K 5/00, 14/00, 16/00; C07H 21/02, 21/04

A. CLASSIFICATION OF SUBJECT MATTER: US CL :

435/6, 7.1, 91.1, 91.2, 240.2, 252.3; 530/300, 350, 387.1, 388.1; 536/22.1, 23.1, 24.3, 24.32, 24.33

### **B. FIELDS SEARCHED**

Electronic data bases consulted (Name of data base and where practicable terms used):

APS. MEDLINE, BIOSIS, CAPLUS, WPIDS search terms: mycobacter?, tubercul?, bovis?, BCG, calmette, guerin, DNA, RNA, oligo, nucleic, oligonucleotide, hybridi?, probe, primer, amplis?, PCR, polymerase chain, ligase chain, LCR, attenuat?, immunoassay, antibod?, monoclon?, polyclon?, protein, peptide, antigen, virulenc?, infect?



# CORRECTED VERSION\*

### **PCT**

### WORLD INTELLECTUAL PROPERTY ORGANIZATION International Bureau



### INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6:

C12Q 1/68, G01N 33/53, C12P 19/34, C12N 5/10, 1/21, C07K 5/00, 14/00, 16/00, C07H 21/02, 21/04

A1

(11) International Publication Number:

WO 96/25519

(43) International Publication Date:

22 August 1996 (22.08.96)

(21) International Application Number:

PCT/US96/01938

(22) International Filing Date:

15 February 1996 (15.02.96)

(30) Priority Data:

08/390,878

17 February 1995 (17.02.95) US

(71) Applicant: PATHOGENESIS CORPORATION [US/US]; Suite 150, 201 Elliott Avenue West, Seattle, WA 98119 (US).

(72) Inventors: STOVER, Charles, Kendall; 7640 81st Place S.E., Mercer Island, WA 98040 (US). MAHAIRAS, Gregory, G.; 3312 39th West, Seattle, WA 98199 (US).

(74) Agents: HUNTER, Tom et al.; Townsend and Townsend and Crew, Steuart Street Tower, One Market, San Francisco, CA 94105-1492 (US). (81) Designated States: AL, AM, AT, AU, AZ, BB, BG, BR, BY, CA, CH, CN, CZ, DE, DK, EE, ES, FI, GB, GE, HU, IS, JP, KE, KG, KP, KR, KZ, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, TJ, TM, TR, TT, UA, UG, UZ, VN, ARIPO patent (KE, LS, MW, SD, SZ, UG), Eurasian patent (AZ, BY, KG, KZ, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).

### **Published**

With international search report.

(54) Title: VIRULENCE-ATTENUATING GENETIC DELETIONS

(57) Abstract

The present invention provides specific genetic deletions that result in an avirulent phenotype of a mycobacterium. These deletions may be used as phenotypic markers of providing a means for distinguishing between disease-producing and non-disease producing mycobacteria.

### FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AM	Armenia	GB	United Kingdom	MW	Malawi
AT	Austria	GE	Georgia	MX	Mexico
AU	Australia	GN	Guinea	NE	Niger
BB	Barbados	GR	Greece	NL	Netherlands
BE	Belgium	HU	Hungary	NO	Norway
BF	Burkina Faso	ΙE	Ireland	NZ	New Zealand
BG	Bulgaria	IT	Italy	PL	Poland
BJ	Benin	JP	Japan	PT	Portugal
BR	Brazil	KE	Kenya	RO	Romania
BY	Belarus	KG	Kyrgystan	RU	Russian Federation
CA	Canada	KP	Democratic People's Republic	SD	Sudan
CF	Central African Republic		of Korea	SE	Sweden
CG	Congo	KR	Republic of Korea	SG	Singapore
CH	Switzerland	KZ	Kazakhstan	SI	Slovenia
CI	Côte d'Ivoire	Ll	Liechtenstein	SK	Slovakia
CM	Cameroon	LK	Sri Lanka	SN	Senegal
CN	China	LR	Liberia	SZ	Swaziland
CS	Czechoslovakia	LT	Lithuania	TD	Chad
CZ	Czech Republic	LU	Luxembourg	TG	Togo
DE	Germany	LV	Latvia	TJ	Tajikistan
DK	Denmark	MC	Monaco	TT	Trinidad and Tobago
EE	Estonia	MD	Republic of Moldova	UA	Ukraine
ES	Spain	MG	Madagascar	UG	Uganda
FI	Finland	ML	Mali	US	United States of America
FR	France	MN	Mongolia	UZ	Uzbekistan
GA	Gabon	MR	Mauritania	VN	Viet Nam

10

15

20

25

30

# VIRULENCE-ATTENUATING GENETIC DELETIONS

### BACKGROUND OF THE INVENTION

Mycobacterium tuberculosis (MTB) infects over ten million people each year and kills over three million, making it the infectious agent causing the greatest mortality worldwide. In an effort to combat Mycobacterium tuberculosis, vaccination programs using a viable attenuated strain of Mycobacterium bovis called bacille Calmette-Guérin (BCG) have been established in more than 120 countries over the course of the last 5 decades. Although widely used and considered safe enough to administer to infants, the BCG vaccine is controversial for two principle reasons: 1) Efficacy for BCG vaccines against tuberculosis has varied from 0-85% in different clinical trials; and 2) Immunization with BCG sensitizes vaccinees to the tubercular antigens used in the tuberculin skin test, confounding attempts to discriminate between BCG immunization and TB infection. For these two reasons, especially the latter, BCG is not used in the United States where surveillance with the tuberculin test is preferred.

The original Pasteur BCG strain was developed by multiple (230 times) serial passages in liquid culture. BCG has never been shown to revert to virulence in animals indicating that the attenuating mutations in BCG are stable deletions and/or multiple mutations which cannot revert. However, the mutations which arose during serial passage of the original BCG strain have never been identified. Moreover, recent efforts to genetically complement BCG virulence with genomic libraries of virulent tubercle bacilli have also been unsuccessful again suggesting that multiple unlinked mutations are responsible for the attenuation of BCG virulence. The antigenicity of BCG and the characteristics leading to its avirulence are thus poorly understood.

### SUMMARY OF THE INVENTION

The present invention provides specific genetic deletions that account for the avirulent phenotype of the bacille Calmette-Guérin (BCG) strain of *Mycobacterium bovis*. These deletions may be used as phenotypic markers of providing a means for distinguishing between disease-producing and non-disease producing mycobacteria.

10

15

20

25

In a preferred embodiment, this invention provides for nucleic acid sequences that are markers for avirulent or virulent mycobacteria. The sequences uniquely characterize the presence or absence of deletions that result in an avirulent phenotype. More specifically the sequence are either deletion junction sequence or deletion sequences or subsequences within deletion junction sequences or deletion sequences. Thus, this invention provides for a marker for an avirulent mycobacterium comprising a first nucleic acid that hybridizes under stringent conditions with a second nucleic acid or a complement of the second nucleic acid where the second nucleic acid or its complement includes BCGala, BCGalb, BCGa2a, BCGa2b, BCGa3a, BCGa3b, BCGa1ab, BCGa2ab, BCGa3ab, BCGa1, BCGa2, and BCGa3. In a particularly preferred embodiment, the marker specifically hybridizes under stringent conditions to a nucleic acid from BCG, but not to a nucleic acid from Mycobacterium tuberculosis or Mycobacterium bovis, or alternatively, the marker specifically hybridizes under stringent conditions to a nucleic acid from Mycobacterium tuberculosis or Mycobacterium bovis, but not to a nucleic acid from BCG. The marker may be the full length BCGa1a, BCGa1b, BCGa2a, BCGa2b, BCGa3a, BCGa3b, BCGa1ab, BCGa2ab, BCGa3ab, BCGa1, BCGa2, and BCGa3 or a subsequence within any of these regions. The marker may also include a nucleic acid having at least 80%, preferably 90%, more preferably 95%, and most preferably 98% percent sequence identity with BCGala, BCGa1b, BCGa2a, BCGa2b, BCGa3a, BCGa3b, BCGa1ab, BCGa2ab, BCGa3ab, BCGa1, BCGa2, or BCGa3. The marker may also include a sequence selected from an open reading frame of a the deletion sequences BCGa1, BCGa2, BCGa3. Suitable open reading frames are indicated in Figures 4, 5, and 6.

The above described marker may be a probe. The probe may be labeled by a number of means including, but not limited to radioactive, fluorescent, enzymatic, and colorimetric labels.

In another embodiment, this invention provides for polypeptides encoded by a subsequence of the BCGa1, BCGa2, or BCGa3 deletions. In particular, the subsequence may be selected from an open reading frame (ORF) present in one of these deletion sequences. This invention also provides for monoclonal or polyclonal antibodies that

10

15

20

25

**\*** 30

specifically bind polypeptides encoded by one or more subsequences of the BCGa1, BCGa2, or BCGa3 deletions.

In still another embodiment, this invention provides for a recombinant cell comprising a first nucleic acid that hybridizes under stringent conditions with a second nucleic acid or a complement of the second nucleic acid where the second nucleic acid or its complement is BCGA1a, BCGA1b, BCGA2a, BCGA2b, BCGA3a, BCGA3b, BCGA1ab, BCGA2ab, BCGA3ab, BCGA1ab, BCGA2ab, BCGA3ab, BCGA1, BCGA2, or BCGA3. The recombinant cell may be a mycobacterium. The recombinant cell may express a polypeptide encoded by any of BCGA1a, BCGA1b, BCGA2a, BCGA2b, BCGA3a, BCGA3b, BCGA1ab, BCGA2ab, BCGA3ab, BCGA1b, BCGA2, and BCGA3. More preferably, the recombinant cell expresses a polypeptide encoded by an intact open reading frame present in any of these regions. The cell may also be a mycobacterium having one or more deletions in the BCGA1, BCGA2, or BCGA3 genomic regions where the deletions result in the attenuation of an otherwise virulent strain of mycobacterium and wherein the deletions are present in up to two of the genomic regions.

In still yet another embodiment, this invention provides a method of distinguishing between an attenuated and a virulent mycobacterium. The method involves detecting the presence or absence of a first nucleic acid that hybridizes under stringent conditions with a second nucleic acid or a complement of the second nucleic acid where the second nucleic acid or its complement is BCGAla, BCGAlb, BCGA2a, BCGA2b, BCGA3a, BCGA3b, BCGAlab, BCGA2ab, BCGA3ab, BCGA1b, BCGA2a, or BCGA3. The first nucleic acid may include any of the markers described above. A particularly preferred marker is one that specifically hybridizes under stringent conditions to a nucleic acid from BCG, but not to a nucleic acid from Mycobacterium tuberculosis or Mycobacterium bovis, or alternatively, that specifically hybridizes under stringent conditions to a nucleic acid from Mycobacterium tuberculosis or Mycobacterium bovis, but not to a nucleic acid from BCG. The method may involve amplifying either the first nucleic acid by any of a number of methods including, for example, polymerase chain reaction. The detection may involve detecting the first nucleic acid, for example, as in a Southern blot, or alternatively, detecting a polypeptide encoded by the first nucleic acid. More specifically, the polypeptide may be

10

15

20

25

30

a encoded by an open reading frame (ORF) selected from BCGa1, BCGa2, or BCGa3. The polypeptide may be visualized by a number of means well known to those of skill in the art including antibody hybridization such as direct or indirect binding of labeled antibody.

This invention additionally provides a method for determining whether an attenuated or a virulent Mycobacterium is present in a sample. This method involves providing a first nucleic acid that hybridizes under stringent conditions with a second nucleic acid or a complement of the second nucleic acid where the second nucleic acid or its complement is BCGala, BCGalb, BCGa2a, BCGa2b, BCGa3a, BCGa3b, BCGalab, BCGa2ab, BCGa3ab, BCGa1, BCGa2, or BCGa3; and hybridizing the first nucleic acid to the biological sample. The first nucleic acid may include any of the markers described above. A particularly preferred marker is one that specifically hybridizes under stringent conditions to a nucleic acid from BCG, but not to a nucleic acid from Mycobacterium tuberculosis or Mycobacterium bovis, or alternatively, that specifically hybridizes under stringent conditions to a nucleic acid from Mycobacterium tuberculosis or Mycobacterium bovis, but not to a nucleic acid from BCG. The method may involve amplifying either the first nucleic acid by any of a number of methods including, for example, polymerase chain reaction. The detection may involve detecting the first nucleic acid, for example, as in a Southern blot, or alternatively, detecting a polypeptide encoded by the first nucleic acid. More specifically, the polypeptide may be a encoded by an open reading frame (ORF) selected from BCGa1, BCGa2, or BCGa3. The method may also include detecting the hybridized first nucleic acid. This may involve direct detection of a label or additionally involve an amplification step and subsequent detection of the amplified product.

Finally, this invention provides a method of producing an attenuated-virulence mycobacterium. This method involves deleting from the genomic DNA of a virulent mycobacterium a first nucleic acid that specifically hybridizes under stringent conditions with a second nucleic acid or a complement of said second nucleic acid where said second nucleic acid or complement of said second nucleic acid is selected from the group consisting of BCGA1, BCGA2, and BCGA3. The first nucleic acid may be BCGA1, BCGA2, or BCGA3, or alternatively, it may be a promoter, other control element or an open reading frame from BCGA1, BCGA2, or BCGA3.

#### **Definitions**

5

10

15

20

25

30

Although any methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present invention, the preferred methods and materials are described. For purposes of the present invention, the following terms are defined below.

The phrase "specifically detect" as used herein refers to the process of determining that a particular subsequence is present in a DNA sample. A DNA sequence may be specifically detected through a number of means known to those of skill in the art. These would include, but are not limited to amplification of the particular target sequence through polymerase chain reaction or ligase chain reaction, hybridization of the sequence to a labeled probe, and binding by labelled ligands or monoclonal antibodies. For a discussion of various means of detection of specific nucleic acid sequences see Perbal, B. A Practical Guide to Molecular Cloning, 2nd Ed. John Wiley & Sons, N.Y. (1988) which is incorporated herein by reference.

The phrase "select subsequence" is used herein to refer to a particular DNA subsequence that is of interest. It is often a predetermined or known sequence of nucleic acid bases. A select subsequence is typically chosen because of a unique sequence identity. Typically a select subsequence is targeted for DNA amplification and often is useful as a specific marker for the presence of a particular gene or a deletion of a particular nucleic acid sequence.

The term "oligonucleotide" refers to a molecule comprised of two or more deoxyribonucleotides or ribonucleotides. Oligonucleotides may include, but are not limited to, primers, probes, nucleic acid fragments to be detected, and nucleic acid controls. Oligonucleotides include naturally occurring nucleotides, chemically modified naturally occurring nucleotides and synthetic nucleotides. The exact size of an oligonucleotide depends on many factors and the ultimate function or use of the oligonucleotide.

The term "primer" refers to an oligonucleotide, whether natural or synthetic, capable of acting as a point of initiation of DNA synthesis under conditions in which synthesis of a primer extension product complementary to a nucleic acid strand is induced, i.e., in the presence of four different nucleoside triphosphates and an agent for polymerization (i.e., DNA polymerase or reverse transcriptase) in an appropriate buffer and at a suitable temperature. A primer is preferably a single-stranded oligodeoxyribonucleotide.

10

15

20

25

30

The appropriate length of a primer depends on the intended use of the primer but typically ranges from 15 to 25 nucleotides. Short primer molecules generally require cooler temperatures to form sufficiently stable hybrid complexes with the template. A primer need not reflect the exact sequence of the template but must be sufficiently complementary to hybridize with a template.

The phrase "PCR primers competent to amplify" as used herein refers to a pair of PCR primers whose sequences are complementary to DNA subsequences immediately flanking the DNA subsequence (target sequence) which it is desired to amplify. The primers are chosen to bind specifically those particular flanking subsequences and no other sequences present in the sample. The PCR primers are thus preferably chosen to amplify the unique target sequence and no other. Alternatively, the PCR primers may be selected to bind to sequences other than the target sequence where the amplification products can be subsequently distinguished (e.g. where the desired amplified sequence is different in size than other amplified sequences).

"Amplifying" or "amplification", which typically refer to an "exponential" increase in target nucleic acid, are used herein to describe both linear and exponential increases in the number of a select target sequence of nucleic acid.

The term "antisense orientation" refers to the orientation of nucleic acid sequence from a structural gene that is inserted in an expression cassette in an inverted manner with respect to its naturally occurring orientation. When the sequence is double stranded, the strand that is the template strand in the naturally occurring orientation becomes the coding strand, and vice versa.

The term "deletion" refers to a region of a nucleic acid which is not present in an organism, but which is present in another related organism. In the context of mycobacteria, a deletion refers, e.g., to a region of nucleic acid which is not present in one strain of mycobacteria, but which is present in another related strain. For instance, an avirulent mycobacterial strain can have a deletion in its genome relative to the genome of a related virulent mycobacterial strain.

The term "deletion junction" refers to the region of a nucleic acid spanning the insertion point of a deletion. Thus, where a region of a nucleic acid sequence is deleted (i.e. a deletion is present), the deletion junction spans the nucleotides that are immediately adjacent to the deletion. Conversely, where a region of a nucleic acid sequence is not

WO 96/25519 7

deleted (i.e. the deletion is absent), two deletion junctions are present, each spanning respectively one end of the deletion sequence and its flanking sequence.

PCT/US96/01938

The following terms are used to describe the sequence relationships between two or more polynucleotides: "reference sequence", "comparison window", "sequence identity", "percentage of sequence identity", and "substantial identity". A "reference sequence" is a defined sequence used as a basis for a sequence comparison; a reference sequence may be a subset of a larger sequence, for example, as a segment of a full-length cDNA or gene sequence given in a sequence listing, such as a polynucleotide sequence of Figures 1, 2, or 3, or may comprise a complete cDNA or gene sequence.

10

15

5

Generally, a reference sequence is at least 10 nucleotides in length, frequently at least 20 to 25 nucleotides in length, and often at least 50 nucleotides in length. Sequence comparisons between two (or more) polynucleotides are typically performed by comparing sequences of the two polynucleotides over a "comparison window" to identify and compare local regions of sequence similarity. A "comparison window", as used herein, refers to a segment of at least 10 contiguous nucleotide positions wherein a polynucleotide sequence may be compared to a reference sequence of at least 10 contiguous nucleotides and wherein the portion of the polynucleotide sequence in the comparison window may comprise additions or deletions (i.e., gaps) of 20 percent or less as compared to the reference sequence (which does not comprise additions or deletions) for optimal alignment of the two sequences.

20

25

Optimal alignment of sequences for aligning a comparison window may be conducted by the local homology algorithm of Smith and Waterman Adv. Appl. Math. 2: 482 (1981), by the homology alignment algorithm of Needleman and Wunsch J. Mol. Biol. 48: 443 (1970); by the search for similarity method of Pearson and Lipman Proc. Natl. Acad. Sci. (USA) 85: 2444 (1988), or by computerized implementations of these algorithms (GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package Release 7.0, Genetics Computer Group, 575 Science Dr., Madison, WI), or by inspection, and the best alignment (i.e., resulting in the highest percentage of sequence similarity over the comparison window) generated by the various methods is selected.

30

The term "sequence identity" means that two polynucleotide sequences are identical (i.e., on a nucleotide-by-nucleotide basis) over the window of comparison. The term "percentage of sequence identity" is calculated by comparing two optimally aligned

10

15

20

25

30

sequences over the window of comparison, determining the number of positions at which the identical nucleic acid base (e.g., A, T, C, G, U, or I) occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the window of comparison (i.e., the window size), and multiplying the result by 100 to yield the percentage of sequence identity. The term "identical" in the context of two nucleic acid or polypeptide sequences refers to the residues in the two sequences which are the same when aligned for maximum correspondence.

The terms "isolated" or "biologically pure" refer to material which is substantially or essentially free from components which normally accompany it as found in its native state. The isolated nucleic acid probes of this invention do not contain materials normally associated with their *in situ* environment, in particular nuclear, cytosolic or membrane associated proteins or nucleic acids other than those nucleic acids intended to comprise the nucleic acid probe itself.

The term "marker" refers to a characteristic which distinguishes one class of cells or compositions from a second class of cells or compositions. For instance, the deletions and deletion junctions described herein can be used to distinguish between strains (e.g., virulent and avirulent strains) of mycobacteria. While markers are indicators of associated features or properties, as used herein, markers may also be used for purposes other than indicating the associated feature or property. Thus, for example, a nucleic acid marker of virulence identifies a particular nucleic acid which may be used in a variety of contexts other than simply indicating virulence.

The term "nucleic acid" refers to a deoxyribonucleotide or ribonucleotide polymer in either single- or double-stranded form, and unless otherwise limited, encompassing known analogues of natural nucleotides that can function in a similar manner as naturally occurring nucleotides. Unless otherwise indicated, a particular nucleic acid sequence includes the complementary sequence thereof.

The term "operably linked" refers to functional linkage between a promoter and a second sequence, wherein the promoter sequence initiates transcription of RNA corresponding to the second sequence.

The term "peptide" or "polypeptide" refers to an amino acid polymer which is encoded by a nucleic acid. The peptide or polypeptide may include naturally occurring or modified amino acids.

WO 96/25519

5

10

15

20

25

30

The terms "probe" or "nucleic acid probe" refer to a molecule that binds to a specific sequence or subsequence of a nucleic acid. A probe is preferably a nucleic acid which binds through complementary base pairing to the full sequence or to a subsequence of a target nucleic acid. It will be understood by one of skill in the art that probes may bind target sequences lacking complete complementarily with the probe sequence depending upon the stringency of the hybridization conditions. The probes are preferably directly labelled as with isotopes, chromophores, lumiphores, chromogens, or indirectly labelled such with, e.g., biotin to which a streptavidin complex may later bind. By assaying for the presence or absence of the probe, one can detect the presence or absence of the selected sequence or subsequence.

The term "labeled nucleic acid probe" refers to a nucleic acid probe that is bound, either covalently, through a linker, or through ionic, van der Waals or hydrogen "bonds" to a label such that the presence of the probe may be detected by detecting the presence of the label bound to the probe.

The term "recombinant" when used with reference to a cell indicates that the cell replicates or expresses a nucleic acid, or expresses a peptide or protein encoded by DNA whose origin is exogenous to the cell. Recombinant cells can express genes that are not found within the native (non-recombinant) form of the cell. Recombinant cells can also express genes found in the native form of the cell wherein the genes are re-introduced into the cell by artificial means.

The term "sample" refers to a material with which bacteria may be associated. Frequently the sample will be a "clinical sample" which is a sample derived from a patient. Such samples include, but are not limited to, sputum, blood, blood cells (e.g., white cells), tissue or fine needle biopsy samples, urine, peritoneal fluid, and pleural fluid, or cells therefrom. It will be recognized that the term "sample" also includes supernatant from eukaryotic cell cultures (which may contain free bacteria), cells from cell or tissue culture, and other media in which it may be desirable to detect mycobacteria (e.g., food and water).

The term "subsequence" in the context of a particular nucleic acid sequence refers to a region of the nucleic acid equal to or smaller than the specified nucleic acid.

The term "substantial identity" or "substantial similarity" indicates that a nucleic acid or polypeptide comprises a sequence that has at least 90% sequence identity to a reference sequence, or preferably 95%, or more preferably 98% sequence identity to the

10

15

20

25

30

reference sequence, over a comparison window of at least about 10 to about 100 nucleotides or amino acid residues. An indication that two polypeptide sequences are substantially identical is that one protein is immunologically reactive with antibodies raised against the second protein. An indication that two nucleic acid sequences are substantially identical is that the polypeptides which the first nucleic acids encodes is immunologically cross reactive with the polypeptide encoded by the second nucleic acid.

Another indication that two nucleic acid sequences are substantially identical is that the two molecules hybridize to each other under stringent conditions. Stringent conditions are sequence-dependent and will be different with different environmental parameters. Generally, stringent conditions are selected to be about 5°C to 20°C lower than the thermal melting point ( $T_m$ ) for the specific sequence at a defined ionic strength and pH. The  $T_m$  is the temperature (under defined ionic strength and pH) at which 50% of the target sequence hybridizes to a perfectly matched probe. Typically, stringent conditions will be those in which the salt concentration is at least about 0.2 molar at pH 7 and the temperature is at least about 60°C.

The term "uninterrupted reading frame" or "open reading frame" refers to a DNA sequence (e.g., cDNA) lacking a stop codon or other intervening, untranslated sequence. An intact open reading frame refers to a full length uninterrupted reading frame or minor variations thereof.

The term "virulent" in the context of mycobacteria refers to a bacterium or strain of bacteria that replicates within a host cell or animal at a rate that is detrimental to the cell or animal within its host range. More particularly virulent mycobacteria persist longer in a host than avirulent mycobacteria. Virulent mycobacteria are typically disease producing and infection leads to various disease states including fulminant disease in the lung, disseminated systemic milliary tuberculosis, tuberculosis meningitis, and tuberculosis abscesses of various tissues. Infection by virulent mycobacteria often results in death of the host organism. Typically, infection of guinea pigs is used as an assay for mycobacterial virulence. In contrast, the term "avirulent" refers to a bacterium or strain of bacteria that either does not replicate within a host cell or animal within its host range, or replicates at a rate that is not significantly detrimental to the cell or animal.

The term BCG-like avirulence, as used herein refers to an attenuated virulence brought about by one of the deletions of the present invention.

10

20

25

ັ 30

### BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 shows the complete sequence listing of the BCG deletion region 1 including flanking sequences. The deletion, designated BCG \( \Delta 1 \), is located between nucleotide 2327 and nucleotide 11126.

Figure 2 shows the complete sequence listing of the BCG deletion region 2 including flanking sequences. The deletion, designated BCG \( \Delta 2 \), is located between nucleotide 3382 and nucleotide 14071.

Figure 3 shows the complete sequence listing of the BCG deletion region 3 including flanking sequences. The deletion, designated BCG \( \Delta \)3, is located between nucleotide 1406 and nucleotide 10673. "N" represents "A", "C", "G", or "T".

Figure 4 shows a map of the deletion sequence BCGa1. This map identifies the various open reading frames (ORFs) and indicates their location within the deletion sequence. Ribozome binding sites and homologies to the predicted encoded proteins are shown.

Figure 5 shows a map of the deletion sequence BCGa2. This map identifies the various open reading frames (ORFs) and indicates their location within the deletion sequence. Ribozomal binding sites and homologies to the predicted encoded proteins are shown.

Figure 6 shows a map of the deletion sequence BCGa3. This map identifies the various open reading frames (ORFs) and indicates their location within the deletion sequence. Ribozome binding sites and homologies to the predicted encoded proteins are shown. The sequence of a small region, estimated to be much less than 200 bp and located close to 9400 bp in Figure 3, remains to be determined. Therefore, the base pair coordinates given in the region 3 map 3' to the 9kb marker are approximations. The precise sequence determination of this region is likely to effect the length of open reading frames 3H and 3L.

Figure 7 illustrates the deletion junction regions of BCGa1, BCGa2, and BCGa3. The "terminal" deletion junction regions formed by the flanking sequences and the terminal regions of the deletion sequences are identified as BCGa1a, BCGa1b, BCGa2a, BCGa2b, and BCGa3a, and BCGa3b. When the deletion is present (the deletion sequences

10

15

20

25

30

are missing) the respective "a" and "b" sequences will be juxtaposed, thereby forming deletion "spanning" junction sequences designated BCG $\triangle$ 1ab, BCG $\triangle$ 2ab, and BCG $\triangle$ 3ab, respectively.

Figure 8 shows EcoRI and BamHI restricted chromosomal DNAs from Mycobacterium bovis, BCG Connaught, and Mycobacterium tuberculosis strains H37Ra, H37Rv, and Erdman probed with <sup>32</sup>P labeled BCG subtracted probe.

### DETAILED DESCRIPTION

This invention reflects the discovery of genetic deletions in mycobacteria that result in an avirulent genotype such as is exhibited by the bacille Calmette-Guérin (BCG) mycobacterium. The original Pasteur bacille Calmette-Guérin (BCG) strain was developed by multiple (230 times) serial passages in liquid culture. BCG has never been shown to revert to virulence in animals indicating that the attenuating mutations in BCG are stable deletions and/or multiple mutations that cannot revert. The mutations that arose during serial passage of the original BCG strain were not previously known. Recent efforts to genetically complement BCG virulence with genomic libraries of virulent tubercle bacilli were unsuccessful, again suggesting that multiple unlinked mutations are responsible for the attenuation of BCG virulence.

The genetic deletions leading to the avirulent phenotype of BCG were identified by genomic subtractions between Connaught strain of BCG and MBV/MTB. The subtracted probe resulting from the genomic subtraction between BCG and the H37 Rv strain of M. tuberculosis was subsequently used to identify and clone three regions from a cosmid library of Mycobacterium bovis genomic DNA. Southern blot mapping and DNA sequence comparisons between BCG and M. bovis showed that three regions, designated regions 1-3, contained DNA segments of approximately 9 kb, 11 kb and 9 kb respectively, which are deleted in the Connaught strain of BCG. Precise deletion junctions were identified for each region by comparisons of BCG and corresponding virulent MBV sequences. The respective deletions, designated BCGa1, BCGa2 and BCGa3 are illustrated in Figures 1-3.

One of skill in the art will appreciate that the deletions encompassed by BCGA1, BCGA2 and BCGA3 may be utilized in a variety of contexts. For example, the deletions may be utilized to distinguish between avirulent and virulent strains of

10

15

20

25

30

mycobacteria thereby providing early detection of patients at risk for tuberculosis. This is of particular importance where mycobacteria are identified in a sample from a patient that has been previously vaccinated with BCG. In this context it may be critical to determine whether mycobacteria identified in a biological sample from such a patient are pathogenic.

In another embodiment, the preparation of mycobacteria containing the deletions of the present invention may provide superior vaccines to BCG which has long been known to have marginal efficacy. Thus, for example, a *Mycobacterium tuberculosis* may contain a full BCGal deletion or a smaller deletion within BCGal (e.g. one or more open reading frames) rendering it avirulent. An avirulent MTB will provide a more efficient vaccine because it is antigenically more similar to MTB than is BCG. Moreover, an MTB rendered avirulent by the production of smaller deletions within the deletion regions identified in this invention will present more antigenic determinants.

Since the loss of virulence is due to the loss of gene products expressed by the nucleic acid sequences comprising the deletion regions, the BCGA1, BCGA2 and BCGA3 deletion sequences and proteins encoded within these deletion sequences provide suitable targets for drug screening. Thus, the use of deleted sequences as targets to screen for drugs that inhibit or interfere with transcription, translation, or post-translational processing of proteins encoded by the deletion sequences, or with the deletion encoded polypeptides themselves, provides an assay for anti-mycobacterial agents. In particular, the use of reporter genes such as firefly luciferase (FFlux), \( \beta\)-galactosidase (BGal), and the like, under the control of promoters present in the deletion sequence provide a rapid assay for drugs regulating activity originating in this region. Conversely, since the protein products of the deletion sequences are presumably expressed in virulent mycobacterial species, proteins expressed by deletion sequences may make good antigens for antimycobacterial vaccines.

Finally, as the viability of BCG demonstrates, deletion regions BCGa1, BCGa2 and BCGa3 are not required for mycobacterial growth and reproduction. Thus, these deletion regions provide good insertion points for the expression of heterologous DNA. The heterologous DNA sequences may be under the control of endogenous inducible or constitutive promoters typically found in the deletion sequences, or alternatively, they may be under the control of introduced promoters, either constitutive or inducible, exogenous to mycobacteria.

10

15

20

#### I. Detection of Deletions

As indicated above, the deletions identified in the present invention provide useful markers for the identification of an avirulent (or conversely a virulent) mycobacterial phenotype. Specifically, determination of avirulence simply requires the detection of the presence or absence of the deletion (either BCGa1, BCGa2, or BCGa3, or deletions within these regions). Where the deletion is present in the bacterial DNA, the bacterium expresses a BCG-like avirulent phenotype. Conversely, where the deletion is absent in the bacterial DNA, the bacterium does not express a BCG-like avirulence. While this may indicate that the bacterium is virulent, one of skill will appreciate that the bacterium may still be avirulent due to the presence of other mutations or deletions. Nevertheless, screening for the presence of the deletion provides a means of detecting a BCG-like avirulent mycobacterium.

Means of detecting deletions are well known to those of skill in the art. Generally, the deletions may be detected either by detecting the presence or absence of deletion junctions, or, alternatively, by detecting the presence or absence of the sequences contained within the deletion (deletion sequences). Where a nucleic acid sequence is deleted (i.e., a deletion is present), the sequences that previously flanked the deleted sequence are juxtaposed, thereby forming a new deletion junction that spans the deletion. Detection of the presence of such a "spanning" deletion junction indicates the presence of the deletion and thus the avirulent phenotype.

Conversely, where the nucleic acid sequence is not deleted (the deletion is not present) the spanning junction sequence will be absent (See, e.g. Figure 7). The "terminal" deletion junction sequences flanking each endpoint of the deletion region are present and detection of these terminal deletion junctions indicates the absence of a deletion. Spanning deletion junction regions and terminal deletion junctions suitable for detecting the deletions of the present invention are illustrated in Figure 7 and in Table 1.

Table 1. Nucleic acid sequences comprising deletion junctions. The symbol "|" indicates the insertion point of the deletion sequence. Deletion sequence bases are represented in lower case letters.

Junction	Nucleotide Sequence	Seq. ID
BCG∆la	CTGGTCGACGATTGGCACAT   gcagccgtgggtgccgccgg	1

30

25

BCG∆1b	gtgtcttcatcggcttccac   CCAGCCGCCCGGATCCAGCA	2
BCG∆2a	CAACTCCACGGCGACCACCC   gcgcccccgctcgcactaga	3
BCG∆2b	gcccacccggtcgagcaccc   CGATGATCTTCTGTTTGACC	4
BCG∆3a	CACCTCGACCACGGCCAACC   gtggacctgtgagatacact	5
BCG∆3b	tcagcagtccacggccaacc   CCGCACCAACACCTTCCACC	6
BCG△lab	CTGGTCGACGATTGGCACAT CCAGCCGCCCGGATCCAGCA	7
BCG∆2ab	CAACTCCACGGCGACCACCC CGATGATCTTCTGTTTGACC	8
BCG <sub>△</sub> 3ab	CACCTCGACCACGGCCAACC CCGCACCAACACCTTCCACC	9

15

5

Where a deletion is detected by determining the presence or absence of sequences contained within the deletion (deletion sequences), the absence of deletion sequences indicates the presence of a deletion and thus an avirulent phenotype. Conversely, the presence of deletion sequences indicates the absence of a deletion. Deletion sequences that provide suitable targets for detecting the deletions of the present invention are provided in Figures 1, 2 and 3.

## A) Isolation of DNA for Detection of Mycobacterium Genomic Deletions

20

In a preferred embodiment, DNA is obtained from mycobacteria. As used herein, the term "mycobacteria" refers to any bacteria of the family Mycobacteriaceae (order Actinomycetales) and includes, but is not limited to, Mycobacterium tuberculosis, Mycobacterium avium complex, Mycobacterium kansasii, Mycobacterium scrofulaceum, Mycobacterium bovis and Mycobacterium leprae. These species and groups and others are described in Baron, S., ed. Medical Microbiology, 3rd Ed. (1991) Churchill Livingstone, New York, which is incorporated herein by reference.

25

The identification of deletions using a DNA marker requires that the DNA sequence be accessible to the particular probes used or to the components of the amplification system if the DNA sequence is to be amplified. In general, this accessibility is ensured by isolating the nucleic acids from the sample.

30

A variety of techniques for extracting nucleic acids from biological samples are known in the art. For example, see those described by Sambrook et al., Molecular Cloning - A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, New

10

15

20

25

30

York, (1985), by Han, et al. Biochemistry, 26: 1617-1625 (1987) and by Du, et al. Bio/Technology, 10: 176-181 (1992), which are incorporated herein by reference.

Alternatively, if the sample is readily disruptable, the nucleic acid need not be purified prior to amplification by the PCR technique, *i.e.*, if the sample is comprised of cells, particularly peripheral blood lymphocytes or monocytes, lysis and dispersion of the intracellular components may be accomplished merely by suspending the cells in hypotonic buffer or boiling them in a low concentration of alkali (*i.e.* 10 mM NaOH).

In a preferred embodiment, DNA is extracted from mycobacteria as described in Example 1.

### B) Detection of Deletions Using Hybridization Probes

In one embodiment the avirulence deletions are detected by contacting DNA obtained from the mycobacterium with a probe that specifically binds an entire deletion junction region or a subsequence of that region and does not specifically bind to any other DNA sequences in the sample. Alternatively, a probe that specifically binds the entire deleted region or subsequence of that region and does not specifically bind to any other sequences in the sample is also suitable. While such probes may be proteins, oligonucleotide probes are preferred. Typically, the sequence of the oligonucleotide probe is chosen to be complementary to a select subsequence unique to the deletion junction or the deletion sequence, whose presence or absence is to be detected. Under stringent conditions the probe will hybridize with the select subsequence forming a stable duplex.

The probe is typically labeled. Detection of the label in association with the target DNA indicates either the presence or absence of the deletion. The probe may be used to detect the deletion junction or deletion sequences directly in a DNA sample without amplification of the deletion subsequences. In one embodiment, unamplified DNA sequences are probed using a Southern blot. The DNA of the sample is immobilized, on a solid substrate, typically a nitrocellulose filter or a nylon membrane. The substrate-bound DNA is then hybridized with the labeled probe under stringent conditions and non-specifically hybridized probe is washed away. Labeled probe detected in association with the immobilized mycobacterial sequences (e.g. bound to the substrate) indicates the presence of deletion sequences (e.g. BCG $\Delta$ 1, BCG $\Delta$ 2, or BCG $\Delta$ 3) and therefore the absence of the deletion. Means for detecting specific DNA sequences are well known to those of skill in

the art. Protocols for Southern blots as well as other detection methods are provided in Maniatis, et al. Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory Press, NY (1982), which is incorporated herein by reference.

In another embodiment, the mycobacterial DNA subsequences are themselves labeled. They are then hybridized, under stringent conditions, with a probe immobilized on a solid substrate. Detection of the label in association with the immobilized probe indicates the presence or absence of the deletion.

In a preferred embodiment, the deletion junction sequences or subsequences or the deletion sequences or subsequences may be amplified by a variety of DNA amplification techniques (for example via cloning, polymerase chain reaction, ligase chain reaction, transcription amplification, etc.) prior to detection using a probe. Because the copy number of mycobacterial sequences bearing the virulence-attenuating deletions is low, the use of unamplified mycobacterial DNA results in an assay of low sensitivity. Amplification of mycobacterial DNA increases sensitivity of the assay by providing more copies of possible target subsequences. In addition, by using labeled primers in the amplification process, the mycobacterial DNA sequences are labeled as they are amplified.

# C) Selection of Probes for Detection of the Deletion Junction Sequences or the Deletion Sequences

20

25

5

10

15

Full length sequences are provided for the deletions BCGa1, BCGa2, and BCGa3 in Figures 1, 2 and 3 respectively. Using these sequence listings, one of skill in the art may easily determine appropriate probes or primers for the detection of the presence or absence of the deletion junctions or the deletion sequences. Generally speaking, a probe will be selected that hybridizes to the target junction sequences or deletion sequences, but not to other mycobacterial nucleic acid sequences under stringent conditions. The design of hybridization probes is well known in the art. See, for example, Sambrook et al., Molecular Cloning - A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, (1989), which is incorporated herein by reference.

30

In a preferred embodiment, the probe is an oligonucleotide sequence complementary to a subsequence comprising a deletion junction (e.g. BCGΔ1a, BCGΔ1b, BCGΔ2a, BCGΔ2b, BCGΔ3a, BCGΔ3b, BCGΔ1ab, BCGΔ2ab, and BCGΔ3ab) or a

10

15

20

25

30

sequence complementary to a subsequence of a deletion sequence (e.g. BCG $\Delta$ 1, BCG $\Delta$ 2, and BCG $\Delta$ 3). The probe preferably has destabilizing mismatches with subsequences from other regions of the mycobacterial genome.

The exact length of the probe depends on many factors including the length of conserved regions around the deletions, the degree of sequence specificity desired, and the amount of internal complementarity within the probe. Such probes are preferably 17 to 25 bases in length. One of skill will recognize that longer probes specifically hybridize at higher temperatures. Generally, stringent conditions are selected to be about 5°C to 20°C, more preferably about 10°C, lower than the thermal melting point (T<sub>m</sub>) for the specific sequence at a defined ionic strength and pH. Under stringent conditions, the probe will specifically hybridize to a nucleic acid sequence from an avirulent mycobacterium such as BCG, but not to a nucleic acid sequence from a virulent mycobacterium such as MTB or MBV. Alternatively, Under stringent conditions, the probe will specifically hybridize to a nucleic acid sequence from a avirulent mycobacterium such as MTB or MBV, but not to a nucleic acid sequence from a avirulent mycobacterium such as MTB or MBV, but not to a nucleic acid sequence from an avirulent mycobacterium such as BCG.

Oligonucleotide probes can be prepared by any suitable method, including, for example, cloning and restriction of appropriate sequences and direct chemical synthesis by a method such as the phosphotriester method of Narang et al. Meth. Enzymol, 68: 90-99 (1979); the phosphodiester method of Brown et al., Meth. Enzymol. 68:109-151 (1979); the diethylphosphoramidite method of Beaucage et al., Tetra. Lett., 22: 1859-1862 (1981); and the solid support method of U.S. Patent No. 4,458,066.

Probe detectability may be increased by the attachment of a label. As used herein, a label is any composition detectable by spectroscopic, photochemical, biochemical, immunochemical, electrical, optical or chemical means. Useful labels in the present invention include magnetic beads (e.g. Dynabeads<sup>TM</sup>), fluorescent dyes (e.g., fluorescein isothiocyanate, texas red, rhodamine, and the like), radiolabels (e.g., <sup>3</sup>H, <sup>125</sup>I, <sup>35</sup>S, <sup>14</sup>C, or <sup>32</sup>P), enzymes (e.g., horse radish peroxidase, alkaline phosphatase and others commonly used in an ELISA), and colorimetric labels such as colloidal gold or colored glass or plastic (e.g. polystyrene, polypropylene, latex, etc.) beads.

Methods for attaching labels to probes, primers, and antibodies are well known to those of skill in the art. For example, the probe can be labeled at the 5'-end with <sup>32</sup>P by incubating the probe with <sup>32</sup>P-ATP and polynucleotide kinase (see Perbal, A

Practical Guide to Molecular Cloning, 2nd ed. John Wiley, N.Y. (1988)). Other labels may be joined to the probe directly or through linkers. They may be located at the ends of the probe or internally. Methods of attaching labels may be found in Connell, et al., Bio/Techniques 5: 342 (1987), U.S. Patent Nos. 4,914,210, 4,391,904 and 4,962,029, which are incorporated herein by reference. In addition, kits for labelling oligonucleotides are widely available. See, for example, Boehringer Mannheim Biochemicals (Indianapolis, IN) for "Genius" labeling kits based on dioxigenin technology and Clonetech (South San Francisco, CA) for a variety of direct and indirect oligonucleotide labeling reagents.

10

15

20

25

30

5

# D) Detection of Deletions Conferring Avirulence Through Amplification of Unique Subsequences

Deletions are particularly amenable to detection without the use of a hybridization probe. In a preferred embodiment, subsequences are amplified that include a deletion junction. The amplified deletion junction may be a "spanning" deletion junction in which case where the deletion is present (i.e. the deletion sequences are absent), the amplification product is a specific DNA incorporating the deletion junction sequence spanning the deletion (e.g. incorporating flanking sequences from both sides of the deleted sequence). Where the deletion is absent (i.e. deletion sequences are present) and primers are selected so that there are no priming sites within the deletion sequences, amplification is non-existent or alternatively provides a complex mixture of non-specifically amplified fragments. Alternatively, amplification primers may be selected that specifically hybridize to deletion sequences, as long as they are selected to amplify sequences that are distinguishable from the sequence amplified when the deletion is present.

Alternatively, the amplification product may be subsequence of a "terminal" deletion junction in which case absence of the deletion (i.e. the deletion sequences are present) will result in the amplification of the specifically targeted nucleic acid. Conversely, where the deletions are present (i.e. the deletion sequences are absent) there will be no specific amplification of a terminal deletion junction.

Amplification products may be separated by size for characterization. Size separation may be accomplished by a variety of means known to those of skill in the art.

10

15

20

25

30

These methods include, but are not limited to electrophoresis, density gradient centrifugation, liquid chromatography, and capillary electrophoresis. In a preferred embodiment, the fragments are separated by agarose gel electrophoresis. The bands are then stained with a marker to visualize them such as ethidium bromide and the gel is visualized, e.g., using ultraviolet light.

As described above, an agarose gel typically shows 1 band if the deletion is present, reflecting amplification of the deletion-spanning sequence. Where the deletion is absent, amplification results in either no bands, where there are no sequences within the deletion to which the amplification primers may hybridize, or a smear where there is non-specific amplification, or a series of discrete bands distinguishable from the band representing the deletion-spanning sequence where primers are chosen that hybridize to deletion sequences.

# E) Selection of Primers for Amplification of Avirulence Deletions

Amplification of deletion junction sequences or subsequences or deletion sequences or subsequences may be accomplished by methods well known in the art, which include, but are not limited to polymerase chain reaction (PCR) (Innis, et al., PCR Protocols. A guide to Methods and Application. Academic Press, Inc. San Diego, (1990), which is incorporated herein by reference), ligase chain reaction (LCR) (see Wu and Wallace, Genomics, 4: 560 (1989), Landegren, et al., Science, 241: 1077 (1988) and Barringer, et al., Gene, 89: 117 (1990), which are incorporated herein by reference), transcription amplification (see Kwoh, et al., Proc. Natl. Acad. Sci. (U.S.A.), 86: 1173 (1989) which is incorporated herein by reference), and self-sustained sequence replication (see Guatelli, et al., Proc. Nat. Acad. Sci. (U.S.A.), 87: 1874 (1990) which is incorporated herein by reference), each of which provides sufficient amplification so that the target sequence can be detected by nucleic acid hybridization to a probe or by electrophoretic separation. Alternatively, methods that amplify the hybridization probe to detectable levels can be used, such as  $Q\beta$ -replicase amplification. See, for example, Kramer, et al. Nature, 339: 401 (1989), Lizardi, et al. Bio/Technology, 6: 1197 (1988), and Lomell, et al., Clin. Chem. 35: 1826 (1989) which are incorporated herein by reference.

In a preferred embodiment, amplification is by polymerase chain reaction using a pair of primers that flank and thereby amplify a selected deletion junction subsequence. Selection of primers is readily apparent to one of skill in the art using the sequence listings of the present invention. For example, a pair of PCR primers 5'-TCGACGATTGGCACAT-3'  $(T_m=55^{\circ}C)$  and 5'-TCCCTCCTGTATTTGTAT-3'  $(T_m=56^{\circ}C)$  will amplify a 469 base pair sequence including the BCGala deletion junction, while 5'-CGTTCTTCGGAGGTTTC-3'  $(T_m=56^{\circ}C)$  and 5'-GGCGGCTGGGTGGA-3'  $(T_m=60^{\circ}C)$  will amplify a 471 base pair sequence including the BCGalb deletion junction.

10

15

20

25

30

5

# F) Detection of Deletions through Detection of Expression Products of Deletion Sequences

In addition to the detection of deletions by the detection of either the deletion junction sequences or the deletion sequences, one may detect the absence of the deletion by detecting the expression products of the deletion sequences. Thus, for example, where the deletion sequences express a protein, the presence of that protein indicates the absence of the deletion and thus is indicative of a virulent (non BCG-like) phenotype. Such proteins are referred to herein as "deletion polypeptides".

Means of determining proteins expressed by particular nucleic acid sequences are well known to those of skill in the art. Typically this involves determining the longest open reading frame. This may be aided by the identification of initiation sites (e.g. ribozome binding sites). The protein encoded by the largest open reading frame is determined using codon preferences for the specific organism from which the nucleic acid is obtained. The polypeptide sequence listing may then be compared against a sequence database, e.g. GenBank, to determine other sequences sharing substantial sequence identity with the calculated sequence. The expression of the protein may be verified by isolating and then sequencing proteins having the predicted length and charge characteristics.

Once deletion polypeptides are identified they may be detected by routine methods well known to those of skill in the art. Typically this involves isolating and then detecting the polypeptide. The polypeptide may be isolated by a number of means well known to those of skill in the art. This includes typical methods of protein

10

15

20

25

30

purification such as high performance liquid chromatography (HPLC), electrophoresis, capillary electrophoresis, hyperdiffusion chromatography, thin layer chromatography, and the like. Methods of purifying and detecting proteins are well known to those of skill in the art (see, e.g., Methods in Enzymology Vol. 182: Guide to Protein Purification, M. Deutscher, ed. Vol. 182 (1990), which is incorporated herein by reference).

Alternatively, deletion polypeptides sequences may be detected using immunoassays utilizing antibodies specific for the deletion polypeptides. The production of such antibodies and their use in immunoassays is detailed below.

### G) Antibodies to Deletion Polypeptides

Antibodies can be raised to the polypeptides encoded by the nucleic acids corresponding to the open reading frames present in the deletion regions of the present invention (deletion polypeptides). As used herein "antibodies" include immunoglobulin or a population of immunoglobins which specifically bind to an antigen. Thus an antibody may be monoclonal or polyclonal including individual, allelic, strain, or species variants, and fragments thereof, both in their naturally occurring (full-length) forms and in recombinant forms. Additionally, antibodies can be raised to these polypeptides in either their native configurations or in non-native configurations. Anti-idiotypic antibodies may also be used.

#### 1) Antibody Production

A number of immunogens may be used to produce antibodies specifically reactive with deletion polypeptides. Recombinant polypeptides are the preferred immunogen for the production of monoclonal or polyclonal antibodies. Naturally occurring polypeptides may also be used either in pure or impure form. Synthetic peptides made using sequences described herein may also used as immunogens for the production of antibodies.

Recombinant polypeptides are expressed in eukaryotic or prokaryotic cells and purified using standard techniques. The polypeptide is injected into an animal capable of producing antibodies. Either monoclonal or polyclonal antibodies may be generated for subsequent use in immunoassays to measure the presence and quantity of the polypeptide.

10

15

20

25

30

Methods of producing polyclonal antibodies are known to those of skill in the art. In brief, an immunogen, preferably a purified deletion polypeptide is mixed with an adjuvant and animals are immunized with the mixture. The animal's immune response to the immunogen preparation is monitored by taking test bleeds and determining the titer of reactivity to the polypeptide of interest. When appropriately high titers of antibody to the immunogen are obtained, blood is collected from the animal and antisera are prepared. Further fractionation of the antisera to enrich for antibodies reactive to the polypeptide is performed where desired. See, e.g., Coligan (1991) Current Protocols in Immunology Wiley/Greene, NY; and Harlow and Lane (1989) Antibodies: A Laboratory Manual Cold Spring Harbor Press, NY, which are incorporated herein by reference.

Monoclonal antibodies may be obtained by various techniques familiar to those skilled in the art. Description of techniques for preparing such monoclonal antibodies may be found in, e.g., Stites et al. (eds.) Basic and Clinical Immunology (4th ed.) Lange Medical Publications, Los Altos, CA, and references cited therein; Harlow and Lane (1988) Antibodies: A Laboratory Manual CSH Press; Goding (1986) Monoclonal Antibodies: Principles and Practice (2d ed.) Academic Press, New York, NY; and particularly in Kohler and Milstein (1975) Nature 256: 495-497, which discusses one method of generating monoclonal antibodies.

Summarized briefly, this method involves injecting an animal with an immunogen. The animal is then sacrificed and cells taken from its spleen, which are then fused with myeloma cells (See, Kohler and Milstein (1976) Eur. J. Immunol. 6: 511-519, incorporated herein by reference). The result is a hybrid cell or "hybridoma" that is capable of reproducing in vitro.

Colonies arising from single immortalized cells are screened for production of antibodies of the desired specificity and affinity for the antigen, and yield of the monoclonal antibodies produced by such cells is enhanced by various techniques, including injection into the peritoneal cavity of a vertebrate host. Alternatively, one may isolate DNA sequences which encode a monoclonal antibody or a binding fragment thereof by screening a DNA library from human B cells according to the general protocol outlined by Huse et al. (1989) Science 246: 1275-1281. In this manner, the individual antibody species obtained are the products of immortalized and cloned single B

10

15

20

25

30

cells from the immune animal generated in response to a specific site recognized on the immunogenic substance.

Other suitable techniques involve selection of libraries of antibodies in phage or similar vectors. See, Huse et al. Science 246: 1275-1281 (1989); and Ward, et al. Nature 341: 544-546 (1989). The polypeptides and antibodies of the present invention are used with or without modification, including chimeric antibodies. Frequently, the polypeptides and antibodies will be labeled by joining, either covalently or non-covalently, a substance which provides for a detectable signal. A wide variety of labels and conjugation techniques are known and are reported extensively in both the scientific and patent literature. Suitable labels include radionucleotides, enzymes, substrates, cofactors, inhibitors, fluorescent moieties, chemiluminescent moieties, magnetic particles, and the like. Patents, teaching the use of such labels include U.S. Patent Nos. 3,817,837; 3,850,752; 3,939,350; 3,996,345; 4,277,437; 4,275,149; and 4,366,241. Also, recombinant immunoglobulins may be produced. See, Cabilly, U.S. Patent No. 4,816,567; and Queen et al. Proc. Nat'l Acad. Sci. USA 86: 10029-10033 (1989).

Antibodies, including binding fragments and single chain versions, against predetermined fragments of deletion polypeptides can be raised by immunization of animals with conjugates of the fragments with carrier proteins as described above. Monoclonal antibodies are prepared from cells secreting the desired antibody. These antibodies can be screened for binding to normal or defective polypeptides, or screened for agonistic or antagonistic activity, e.g., mediated through a receptor. These monoclonal antibodies will usually bind with at least a  $K_D$  of about 1 mM, more usually at least about 300  $\mu$ M, and most preferably at least about 0.1  $\mu$ M or better.

The antibodies of this invention can also be used for affinity chromatography in isolating deletion polypeptides. Columns can be prepared where the antibodies are linked to a solid support, e.g., particles, such as agarose, Sephadex, or the like, where a bacterial lysate, or recombinant cell lysate is passed through the column, washed, and treated with increasing concentrations of a mild denaturant, whereby purified deletion polypeptides are released.

10

15

20

25

30

The antibodies can be used to screen expression libraries for particular expression products. Usually the antibodies in such a procedure will be labeled with a moiety allowing easy detection of presence of antigen by antibody binding.

In a preferred embodiment, antibodies to deletion polypeptides are used for the identification of cell populations expressing the polypeptides. By assaying the expression products of cells expressing the polypeptides it is possible to diagnose bacterial infections.

Antibodies raised against each polypeptide are useful to raise anti-idiotypic antibodies. These will be useful in detecting or diagnosing various immunological conditions related to the presence of the respective antigens.

### 2) Immunoassays

A particular deletion polypeptide can be measured by a variety of immunoassay methods. For a review of immunological and immunoassay procedures in general, see Stites and Terr (eds.) 1991 Basic and Clinical Immunology (7th ed.). Moreover, the immunoassays of the present invention can be performed in any of several configurations, e.g., those reviewed in Maggio (ed.) (1980) Enzyme Immunoassay CRC Press, Boca Raton, Florida; Tijan (1985) "Practice and Theory of Enzyme Immunoassays," Laboratory Techniques in Biochemistry and Molecular Biology, Elsevier Science Publishers B.V., Amsterdam; and Harlow and Lane Antibodies, A Laboratory Manual, supra, each of which is incorporated herein by reference. See also Chan (ed.) (1987) Immunoassay: A Practical Guide Academic Press, Orlando, FL; Price and Newman (eds.) (1991) Principles and Practice of Immunoassays Stockton Press, NY; and Ngo (ed.) (1988) Non-isotopic Immunoassays Plenum Press, NY.

Immunoassays for measurement of deletion polypeptides can be performed by a variety of methods known to those skilled in the art. In brief, immunoassays to measure the protein can be, e.g., competitive or noncompetitive binding assays. In competitive binding assays, the sample to be analyzed competes with a labeled analyte for specific binding sites on a capture agent bound to a solid surface. Preferably the capture agent is an antibody specifically reactive with a deletion polypeptide produced as described above. The concentration of labeled analyte bound to the capture agent is inversely proportional to the amount of free analyte present in the sample.

10

15

20

25

30

In a competitive binding immunoassay, the deletion polypeptide present in the sample competes with labelled protein for binding to a specific binding agent, for example, an antibody specifically reactive with a particular deletion polypeptide. The binding agent is, e.g., bound to a solid surface to produce separation of bound labelled polypeptide from the unbound labelled polypeptide. Alternately, the competitive binding assay may be conducted in liquid phase and any of a variety of techniques known in the art may be used to separate the bound labelled protein from the unbound labelled protein. Following separation, the amount of bound labeled protein is determined. The amount of polypeptide present in the sample is inversely proportional to the amount of labelled polypeptide binding.

Alternatively, a homogenous immunoassay may be performed in which a separation step is not needed. In these immunoassays, the label on the protein is altered by the binding of the protein to its specific binding agent. This alteration in the labelled protein results in a decrease or increase in the signal emitted by label, so that measurement of the label at the end of the immunoassay allows for detection or quantitation of the polypeptide.

Deletion polypeptides may also be detected by a variety of noncompetitive immunoassay methods. For example, a two-site, solid phase sandwich immunoassay may be used. In this type of assay, a binding agent for the protein, for example an antibody, is attached to a solid support. A second protein binding agent, which is also an antibody, and which binds the protein at a different site, is labelled. After binding at both sites on the protein, the unbound labelled binding agent is removed and the labelled binding agent bound to the solid phase is measured. The amount of labelled binding agent bound is directly proportional to the amount of polypeptide in the sample.

Western blot analysis can be used to determine the presence of a deletion polypeptide in a sample. Electrophoresis is carried out, for example, on a bacterial sample suspected of containing the deletion polypeptide. Following electrophoresis to separate the proteins, and transfer of the proteins to a suitable solid support such as a nitrocellulose filter, the solid support is incubated with an antibody reactive with the protein. This antibody is labelled, or alternatively may be it is detected by subsequent incubation with a second labelled antibody that binds the primary antibody.

10

15

20

25

30

The immunoassay formats described above employ labelled assay components. The label can be in a variety of forms as described above. The choice of label depends on sensitivity required, ease of conjugation with the compound, stability requirements, and available instrumentation. For a review of various labelling or signal producing systems which may be used, see U.S. Patent No. 4,391,904, which is incorporated herein by reference.

Antibodies reactive with a particular protein can also be measured by a variety of immunoassay methods. For a review of immunological and immunoassay procedures applicable to the measurement of antibodies by immunoassay techniques, see Stites and Terr (eds.) Basic and Clinical Immunology (7th ed.) supra; Maggio (ed.) Enzyme Immunoassay, supra; and Harlow and Lane Antibodies, A Laboratory Manual, supra.

In brief, immunoassays to measure antisera reactive with polypeptides include competitive and noncompetitive binding assays. In competitive binding assays, the sample analyte competes with a labeled analyte for specific binding sites on a capture agent bound to a solid surface. Preferably the capture agent is a purified recombinant deletion polypeptide as described above. Other sources of polypeptides, including isolated or partially purified naturally occurring protein, can also be used. Noncompetitive assays are typically sandwich assays, in which the sample analyte is bound between two analyte-specific binding reagents. One of the binding agents is used as a capture agent and is bound to a solid surface. The second binding agent is labelled and is used to measure or detect the resultant complex by visual or instrument means. A number of combinations of capture agent and labelled binding agent can be used. A variety of different immunoassay formats, separation techniques and labels can be also be used similar to those described above for the measurement of deletion polypeptides.

# II. Preparation of Deletion-Containing Mycobacteria

Mycobacteria containing specific deletions may be prepared by using methods of homologous recombination well known to those of skill in the art. In brief, homologous recombination is a natural cellular process which results in the scission of two nucleic acid molecules having identical or substantially similar (i.e. "homologous") sequences, and the ligation of the two molecules such that one region of each initially

10

15

20

25

30

present molecule is now ligated to a region of the other initially present molecule (Sedivy, *Bio/Technol.*, 6: 1192-1196 (1988).

Homologous recombination is exploited by a number of various methods of "gene targeting" well known to those of skill in the art. (see, for example, Mansour et al. Nature, 336: 348-352 (1988); Capecchi Trends Genet. 5: 70-76 (1989); Capecchi Science 244: 1288-1292 (1989); Capecchi et al. pages 45-52 In: Current Communications in Molecular Biology, Capecchi, M.R. (ed.), Cold Spring Harbor Press, Cold Spring Harbor, N.Y. (1989); Frohman et al. Cell 56: 145-147 (1989)). Some approaches focus on increasing the frequency of recombination between two DNA molecules by treating the introduced DNA with agents which stimulate recombination (e.g. trimethylpsoralen, UV light, etc.), however, most approaches utilize various combinations of selectable markers to facilitate isolation of the transformed cells.

One such selection method is termed positive/negative selection (PNS) (Thomas and Cappechi Cell 51: 503-512 (1987)). This method involves the use of two selectable markers: one a positive selection marker such as the bacterial gene for neomycin resistance (neo'); the other a negative selection marker such as the herpes virus thymidine kinase (tk) gene. Neo' confers resistance to the drug G-418, while herpes tk renders cells sensitive to the nucleoside analog gangcyclovir (GANC) or 1-(2-deoxy-2-fluoro-b-d-arabinofuranosyl)-5-iodouracil (FIAU). The DNA encoding the positive selection marker in the transgene (e.g. neo'') is generally linked to an expression regulation sequence that allows for its independent transcription in mycobacteria. It is flanked by first and second sequence portions of at least a part of the deletion or deletion flanking sequences.

These first and second sequence portions target the transgene to a specific nucleotide sequence. A second independent expression unit capable of producing the expression product for a negative selection marker, e.g. for herpes virus tk is positioned adjacent to or in close proximity to the distal end of the first or second portions of the first DNA sequence. Upon transfection, some of the mycobacteria incorporate the transgene by random integration, others by homologous recombination between the endogenous allele and sequences in the transgene. As a result, one copy of the targeted nucleic acid is disrupted by homologous recombination with the-transgene with simultaneous loss of the sequence encoding herpes tk gene. Random integrants, which

10

15

20

25

30

occur via the ends of the transgene, contain herpes tk and remain sensitive to GANC or FIAU. Therefore, selection, either sequentially or simultaneously with G418 and GANC enriches for transfected mycobacteria containing the transgene integrated into the genome by homologous recombination.

Methods of homologous recombination in mycobacteria are described in greater detail by Ganjam et al. Proc. Natl. Acad. Sci. USA, 88: 5433-5437 (1991) and Aldovini et al., J. Bacteriol., 175: 7282-7289 (1993) which are incorporated herein by reference.

## III. Screening for Drug Susceptibility/Therapeutics

The expression products of the open reading frames in the BCGa1, BCGa2, and BCGa3 deletions of the present invention are targets for anti-mycobacterial drugs. To determine particularly suitable drug targets, open reading frames and surrounding expression control sequences are introduced into avirulent strains of mycobacteria, alone or in combination with other open reading frame regions to determine which regions are critical for virulence. Once particular genes are identified as critical for virulence, anti-mycobacterial agents are designed to inhibit expression of the critical genes, or to attack the critical gene products. For instance, antibodies are generated against the critical gene products and used as prophylactic or therapeutic agents. Alternatively, small molecules can be screened for the ability to selectively inhibit expression of the critical gene products, e.g., using recombinant expression systems which include the gene's endogenous promoter. These small molecules are then used as therapeutics, or prophylactic agents to inhibit mycobacterial virulence.

In another embodiment, anti-mycobacterial agents which render a virulent mycobacterium avirulent can be operably linked to expression control sequences and used to transform a virulent mycobacterium. Such anti-mycobacterial agents inhibit the replication of a specified mycobacterium upon transcription or translation of the agent in the mycobacterium.

Such transformed mycobacteria are useful as vaccine components, and as components of immunological infectivity assays. For instance, an animal's blood can be monitored for the presence of anti-mycobacterial antibodies using the procedures described herein, using transformed avirulent mycobacterial components in various

10

15

20

25

30

immunological assays. Anti-mycobacterial agents useful in this invention include, without limitation, antisense genes, ribozymes, decoy genes, transdominant proteins and suicide genes.

An antisense nucleic acid is a nucleic acid that, upon expression, hybridizes to a particular mRNA molecule, to a transcriptional promoter or to the sense strand of a gene. By hybridizing, the antisense nucleic acid interferes with the transcription of a complementary DNA, the translation of an mRNA, or the function of a catalytic RNA. Antisense molecules useful in this invention include those that hybridize to gene transcripts in the region of the deletions of the invention, particularly deletion region 1.

A ribozyme is a catalytic RNA molecule that cleaves other RNA molecules having particular nucleic acid sequences. Ribozymes useful in this invention are those that cleave deletion gene transcripts. Examples include hairpin and hammerhead ribozymes.

A decoy nucleic acid is a nucleic acid having a sequence recognized by a regulatory DNA binding protein (i.e., a transcription factor). Upon expression, the transcription factor binds to the decoy nucleic acid, rather than to its natural target in the genome. Useful decoy nucleic acid sequences include any sequence to which a transcription factor binds in the deletion regions of the present invention.

A transdominant protein is a protein whose phenotype, when supplied by transcomplementation, will overcome the effect of the native form of the protein. For instance, an avirulent mycobacterium can be rendered virulent by introducing transdominant proteins from deletion region 1.

A suicide gene produces a product which is cytotoxic. In the vectors of the present invention, a suicide gene is operably linked to an inducible expression control sequences which is stimulated upon infection of a cell by a mycobacterium.

# IV. Use of Expressed "Deletion Proteins" in a Vaccine

The deletion polypeptides encoded by the open reading frames in BCGa1, BCGa2, and BCGa3 may be recombinantly expressed and used as components of immunological assays as described above or in vaccines. Expression of polypeptides

encoded by the open reading frames of the BCGa1, BCGa2, or BCGa3 deletions may be accomplished by means well known to those of skill in the art.

In brief, the expression of natural or synthetic nucleic acids encoding deletion polypeptides will typically be achieved by operably linking the DNA or cDNA to a promoter (which is either constitutive or inducible), followed by incorporation into an expression vector. The vectors can be suitable for replication and integration in either prokaryotes or eukaryotes. Typical expression vectors contain transcription and translation terminators, initiation sequences, and promoters useful for regulation of the expression of polynucleotide sequence encoding deletion polypeptides.

10

15

5

To obtain high level expression of a cloned gene, such as those polynucleotide sequences encoding deletion polypeptides, it is desirable to construct expression plasmids which contain, at the minimum, a promoter to direct transcription, a ribosome binding site for translational initiation, and a transcription/translation terminator. The expression vectors may also comprise generic expression cassettes containing at least one independent terminator sequence, sequences permitting replication of the plasmid in both eukaryotes and prokaryotes, *i.e.*, shuttle vectors, and selection markers for both prokaryotic and eukaryotic systems. For detailed techniques employed in the recombinant expression of deletion proteins see, for example, Sambrook, et al., Molecular Cloning: A Laboratory Manual (2nd Ed., Vols. 1-3, Cold Spring Harbor Laboratory (1989)), Methods in Enzymology, Vol. 152: Guide to Molecular Cloning Techniques (Berger and Kimmel (eds.), San Diego: Academic Press, Inc. (1987)), or Current Protocols in Molecular Biology, (Ausubel, et al. (eds.), Greene Publishing and Wiley-Interscience, New York (1987), all of which are incorporated herein by reference.

20

30

25 For test whe

For example, the deletion polypeptides can be used as reagents in immunoblot assays to test whether a patient was previously exposed to virulent mycobacteria (i.e., to test whether the patient has antibodies to the deletion polypeptide). These assays have the advantage of discriminating between previous exposure to an avirulent mycobacterium (e.g., one used in a vaccine) and exposure to a virulent mycobacterium. Thus, vaccinated individuals can be tested for antibodies to the virulent mycobacterium without regard to whether the patient has been vaccinated with an avirulent mycobacterium.

The expressed deletion polypeptides may be used in a variety of assays.

10

15

20

25

30

The deletion polypeptides can also be used as antigenic vaccine components to direct antibodies to elements which are critical for virulence. These polypeptides can be added to existing vaccines (e.g., those based upon avirulent mycobacteria and which lack the deletion polypeptide) to supplement the range of antigenicity conferred by the vaccine, or they may be used apart from other mycobacterial antigens. The vaccines of the invention contain as an active ingredient an immunogenically effective amount of a deletion polypeptide or of a recombinant vector which includes the deletion polypeptide. The immune response can include the generation of antibodies; activation of cytotoxic T lymphocytes (CTL) against cells presenting peptides derived from the polypeptides or other mechanisms well known in the art. See e.g. Paul Fundamental Immunology Third Edition published by Raven press New York (incorporated herein by reference) for a description of immune response. Useful carriers are well known in the art, and include, for example, thyroglobulin, albumins such as human serum albumin, tetanus toxoid, and polyamino acids such as poly(D-lysine:D-glutamic acid). The vaccines can also contain a physiologically tolerable (acceptable) diluent such as water, phosphate buffered saline, and further typically include an adjuvant. Adjuvants such as incomplete Freund's adjuvant, aluminum phosphate, aluminum hydroxide, or alum are materials well known in the art.

The compositions are suitable for single administrations or a series of administrations. When given as a series, inoculations subsequent to the initial administration are given to boost the immune response and are typically referred to as booster inoculations.

The vaccine compositions of the invention are intended for parenteral, topical, oral or local administration. Preferably, the pharmaceutical compositions are administered parenterally, e.g., intravenously, subcutaneously, intradermally, or intramuscularly. Thus, the invention provides compositions for parenteral administration that comprise a solution of the agents described above dissolved or suspended in an acceptable carrier, preferably an aqueous carrier. A variety of aqueous carriers may be used, e.g., water, buffered water, 0.4% saline, 0.3% glycine, hyaluronic acid and the like. These compositions may be sterilized by conventional, well known sterilization techniques, or may be sterile filtered. The resulting aqueous solutions may be packaged for use as is, or lyophilized, the lyophilized preparation being combined with a sterile

٠. >

ŧ

Ş.

5

10

15

20

25

30

solution prior to administration. The compositions may contain pharmaceutically acceptable auxiliary substances as required to approximate physiological conditions, such as pH adjusting and buffering agents, tonicity adjusting agents, wetting agents and the like, for example, sodium acetate, sodium lactate, sodium chloride, potassium chloride, calcium chloride, sorbitan monolaurate, triethanolamine oleate, etc.

For solid compositions, conventional nontoxic solid carriers may be used which include, for example, pharmaceutical grades of mannitol, lactose, starch, magnesium stearate, sodium saccharin, talcum, cellulose, glucose, sucrose, magnesium carbonate, and the like. For oral administration, a pharmaceutically acceptable nontoxic composition is formed by incorporating any of the normally employed excipients, such as those carriers previously listed, and generally 10-95% of active ingredient and more preferably at a concentration of 25%-75%.

For aerosol administration, the polypeptides are preferably supplied in finely divided form along with a surfactant and propellant. The surfactant should be nontoxic, and preferably soluble in the propellant. Representative of such agents are the esters or partial esters of fatty acids containing from 6 to 22 carbon atoms, such as caproic, octanoic, lauric, palmitic, stearic, linoleic, linolenic, olesteric and oleic acids with an aliphatic polyhydric alcohol or its cyclic anhydride. Mixed esters, such as mixed or natural glycerides may be employed. A carrier can also be included, as desired, as with, e.g., lecithin for intranasal delivery.

The amount of vaccine administered to the patient will vary depending upon the composition being administered, the physiological state of the patient and the manner of administration.

Live attenuated recombinant viruses which include the deletion polypeptide, such as recombinant vaccinia or adenovirus vectors, are convenient alternatives as vaccines because they are inexpensive to produce and are easily transported and administered. Vaccinia vectors and methods useful in immunization protocols are described, for example, in U.S. Patent No. 4,722,848, incorporated herein by reference.

Deletion sequences and subsequences of this invention may also be used in methods of genetic immunization. Briefly, genetic immunization involves transfecting

10

15

20

25

30

cells in vivo with nucleic acids encoding pathogen specific antigens. The transformed host cells then express the antigen thereby stimulating the host immune system.

In the present invention, antigen-encoding deletion region sequences are used to transform mammalian host cells thereby resulting in the expression of the antigen by the host. This provokes an immune response by the host against the expressed antigen thereby conferring immunity on the host. Methods of genetic immunization are well known to those of skill in the art (see, e.g., Wang et al. Proc. Natl. Acad. Sci. USA, 90: 4156-4160 (1993); Ulmer et al., Science, 259: 1745-1749 (1993); Fynan et al. DNA Cell Biol., 12: 785-789 (1993); Fynan et al. Proc. Natl. Acad. Sci. USA, 90: 11478-11482 (1993); Robinson et al. Vaccine, 11: 957-960 (1993); and Martinon et al. Eur. J. Immunol., 23: 1719-1722 (1993), which are incorporated herein by reference.

# VI. Use of Promoters within Deletion Sequences for Expression of Recombinant Proteins

Bacille Calmette-Guérin (BCG) contains all three deletions (BCGΔ1, BCGΔ2, and BCGΔ3) and yet is able to grow and reproduce indicating that the sequences contained within the deletion are not essential for bacterial viability. These deletion regions therefore make good target sites for the insertion of heterologous DNA as mycobacteria are tolerant of disruption of the native genome in these regions. The BCGΔ1, BCGΔ2, and BCGΔ3 deletion regions therefore provide suitable target sites for the incorporation of expression cassettes and the subsequent expression of exogenous gene products. The expression cassettes typically comprise a nucleic acid sequence under the control of a promoter. The promoter may be either constitutive or inducible. The cassette may additionally comprise a selectable marker such as an antibiotic resistance gene, a gene encoding a fluorescent marker (e.g. green fluorescent protein), or a gene encoding an enzymatic marker (e.g. β-galactosidase).

Alternatively, genes under the control of endogenous promoters may be used as well. In one embodiment, reporter genes under the control of endogenous promoters found within the deletion sequences may be inserted at the deletion sites. These reporter genes may be utilized as an assay for antimycobacterial compounds that act by inhibiting transcription or translation of deletion sequences. Assaying for the

10

15

20

25

30

reporter gene product in the presence of an antimycobacterial compound provides a measure of efficacy of that compound in upregulating or downregulating deletion sequence genes. Methods of use of mycobacterial reporter gene assays to screen for drug activity are described by Cooksey et al., Antimicrob. Agents Chemother., 37: 1348-1352 (1993), and Jacobs et al., Science, 260: 819-822 (1993) which are incorporated herein by reference.

### **EXAMPLES**

The following examples are offered by way of illustration, not by way of limitation.

#### Example 1

## Identification of Virulence-Attenuating Deletions

#### **Bacterial Culture**

All strains of Mycobacteria used in this study were maintained in 7H9 (Difco, Detroit Michigan, USA) media supplemented with OADC (BBL) or were grown on 7H11 agar supplemented with oleic acid albumin dextrose complex (OADC). Escherichia coli (strain DH5 $\alpha$  or NM554) was used as a host for all recombinant plasmids and cosmids. E. coli was maintained in LB medium with or without agar. Carbenicillin (100  $\mu$ g/ml) was used in place of ampicillin for the selection of all E. coli plasmids.

## Extraction of High Molecular Weight DNA

High molecular weight chromosomal DNA was prepared by diluting a late log phase culture of the respective mycobacterium 1:10 into a liter of 7H9 medium containing 1.5% glycine and continuing growth for 4 to 5 days. The cells were then harvested by centrifugation, washed once in TE (pH 8.0) and resuspended in 4 ml of 25% sucrose in 10X TE. 100  $\mu$ g of lysozyme was added and the preparation was incubated at 37°C for 2 hr followed by the addition of 100  $\mu$ g of proteinase K and sarkosyl to a concentration of 1% weight/volume. Following overnight incubation at 65°C the mixture was extracted 4 times with chloroform isoamyl alcohol 24:1, once with phenol/chloroform (1:1), and twice again with chloroform isoamyl alcohol. The resulting high molecular weight DNA was then run on a CsCl gradient as described by

10

15

20

25

30

Hull et al. Infect. Immun., 33: 933-938 (1981), which is incorporated herein by reference, and subsequently dialyzed against 4 changes of TE. BCG DNA was physically sheared by passage through a 22 gauge needle until an average size of 3-10 kb was obtained (20-25 passages). This DNA was then biotinylated using photobiotin (Clonetech, Palo Alto, California, USA) according to the method of Straus and Ausubel, Proc. Natl. Acad. Sci. USA, 87: 1889-1893 (1990), which is incorporated herein by reference.

#### **DNA Subtraction**

DNA subtraction was carried out between virulent *M. tuberculosis* H37Rv and avirulent BCG. H37R chromosomal DNA was selected because it was the most readily available chromosomal DNA from a virulent strain. In addition, *M. bovis* and *M tuberculosis* H37Rv are highly homologous.

M. bovis/M. tuberculosis specific probes were generated by the method of Straus and Ausubel, supra. with the following modifications. Sheared and biotinylated BCG DNA was used in a 10:1 excess for each round of subtraction. Wild type M. tuberculosis H37Rv DNA was digested with Sau3A to an average size of 1 kb. Hybridization conditions were 1M NaCl and 65 °C for 18 hours. Following five cycles (successive denaturation and reassociations) of subtraction, Sau3A1 adaptors (GACACTCTCGAGACATCACCGTCC and GATCGGACGGTGATGTCTCGAGAGTG were ligated to the subtraction product and amplified in a PCR reaction for 35 cycles (30 sec at 95°C, 30 sec at 55°C, and 3 min at 72°C). The M. tuberculosis/M. bovis specific probes were radiolabeled by using one strand of the adaptor (GACACTCTCGAGACATCACCGTCC) as a primer and labeling with <sup>32</sup>P dCTP using the Klenow fragment of DNA polymerase.

An M. bovis cosmid library was constructed in the BamH1 site of sCOS (Stratagene, La Jolla California, USA) with subsequent in vitro packaging and infection of E. coli strain NM554 (Stratagene). 600 colonies were picked to Nytran circular membranes and the membranes prepared according to the method of Grunstein and Hogness, Proc. Natl. Acad. Sci. USA, 72: 3961 (1975), which is incorporated herein by reference. These filters were then probed using the BCG subtracted probe and positive clones selected for further analysis. Cosmid DNA was prepared from selected clones by the method of Birnboim and Doly, Nucleic Acids. Res., 7: 1513 (1973) which is

incorporated herein by reference. Restriction fragments that hybridize with the MTB/MBV specific probe were further subcloned into pGEM7z or pGEM5z (Promega, Madison, Wisconsin, USA) for deletion analysis.

Plasmid DNA for DNA sequencing was prepared using Qiagen minicolumns (Qiagen Inc. Chatsworth California, USA) and sequenced by the method of Henikoff, *Gene*, 28: 351-359 (1984), which is incorporated herein by reference, using the Erase A Base System (Promega). DNA sequencing reactions were run using a Perkin Elmer 9600 thermocycler and analyzed on an automated ABI sequencer. Analysis and assembly of contiguous DNA sequence was done using the ABI analysis software and SeQuencher sequence analysis software by Gene Clones Corp (Ann Arbor, Michigan, USA).

### Deletion Region 1 (BCGA1)

5

10

15

20

25

30

Sequence analysis of over 16 kb of MBV region 1 and homologous regions in BCG revealed the precise junctions for the deletion in BCG. Eight open reading frames were identified with codon usage biases matching that of known MTB and MBV genes (see map Figure 4). The potential start and stop codons and predicted maximum protein coding capacity are listed in Figure 4. Consensus ribosomal binding site sequences were found near potential start codons for seven of eight open reading frames. TBLASTN and FASTA sequence homology analysis with each potential ORF-encoded protein revealed significant homologies for 3 of 8 open reading frames in region 1.

Most notable is the ORF1C homology to an unpublished and uncharacterized sequence listed in Genbank as *M. tuberculosis* antigen esat6. A 65 base pair repeated overlapping (repeated  $\sim 2$  1/2 times) sequence was also recognized within the ORF1C (esat6) open reading frame. Also noteworthy are the significant homologies identified between ORF1H and bacterial serine proteases including *B. subtilus* subtilisin. Of the eight recognized open reading frames, four (ORFs 1B, 1C, 1D, and 1E) are located entirely within the 9 kb region deleted in BCG. One ORF traverses the BCG deletion junction in virulent *M. bovis*.

DNA probes from the 9 kb deletion in region 1 demonstrated that this region is absent in all BCG substrains and present in all virulent MBV and MTB strains tested. Furthermore, restriction fragment patterns observed in Southern blot analysis

10

15

20

25

30

with region 1 probes are non-polymorphic and identical in virulent MBV and MTB. This region has far fewer direct and indirect repeats than the regions 2 (BCG $\Delta$ 2) and 3 (BCG $\Delta$ 3) characterized below.

The sequence of a small region, estimated to be less than 20 bp between basepair coordinates 10654 and 10664 in region 1 has been recalcitrant to automated sequencing. Therefore, pending sequence confirmation, the base pair coordinates given in the region 1 map (Figure 4) are approximations. The precise sequence determination is likely to effect the Orf1E open reading frame.

### Deletion Region 2 (BCGA2)

Sequence analysis of over 15 kb of MBV region 2 and homologous regions in BCG revealed the precise junctions for an 11 kb deletion in BCG. Thirteen open reading frames were identified with codon usage biases matching that of known MTB and MBV genes (see map Figure 5). The potential start and stop codons and predicted maximum protein coding capacity are also shown in Figure 5. Candidate consensus sequences resembling ribosomal binding sites were found near potential start codons for eight open reading frames. Of the thirteen open reading frames recognized in BCGA2, nine are located entirely within the 11 kb region deleted in most BCG strains while ORF2B2 and ORF2I traverse the deletion junctions.

TBLASTN and FASTA sequence homology analysis with each potential ORF-encoded protein revealed significant homologies for five open reading frames in BCG $\Delta$ 2. A protein encoded by ORF2C exhibits striking similarity to the *E. coli* iciA protein which is thought to play a role in inhibiting and regulating the initiation of chromosomal replication. The iciA protein product is a member of the large LysR family of transcriptional regulatory proteins. Orf2F is highly homologous to an *S. typhimurium* ribonucleotide diphosphate reductase and a region of the *E. coli* and *S. typhimurium* proUVWX operon. Orf2H was found to have significant homology to *E. coli* and *S. typhimurium* permeases involved in aromatic amino acid transport and a eukaryotic cell retroviral receptor.

The Orf2G encoded protein was identical to the MTB mpt64 gene previously thought to encode a secreted antigen which is specifically expressed by MTB

39

and not BCG strains. Recent analysis of mpt64 expression revealed that three BCG substrains do express mpt64 (Moreau, Tokyo, Russian). Probes specific for mpt64 or other non-repetitive parts of region 2 hybridized to all MTB strains tested and the same three BCG substrains shown to express mpt64. Of interest is the finding that these three BCG substrains are derived from the original Pasteur strain prior to 1925. The current Pasteur strain and all strains derived from the original Pasteur strain after 1925, including the Connaught strain used in the subtractive analysis in this study, are deleted in the 11 kb DNA segment contained within BCG $\Delta$ 2. These data indicate that an additional mutational event deleting the 11 kb segment of region 2, occurred in the BCG Pasteur strain sometime after 1925.

Southern blot analysis with probes from different segments of region 2 revealed a repetitive element located within a 2 kb segment (8-10 kb) of region 2. This repetitive element is ubiquitous in all tubercle bacilli tested. This element provides a marker suitable for RFLP analysis of mycobacterial strains.

15

20

25

30

10

5

## Deletion Region 3 (BCGA3)

Sequence analysis of the almost 11 kb region 3 sequence and comparison to a homologous region in BCG precisely identified the deletion junctions for BCG. Twelve potential open reading frames were recognized in the region 3 sequence, seven of which are entirely located within the 9 kb region deleted in BCG. At least 9 ORFs in BCG $\Delta$ 3 exhibit codon usage preferences comparable to that of the tubercle bacilli. Sequence homology analysis of presumptive protein sequences encoded by six open reading frames in region 3 revealed highly significant homology to listed sequences. Orfs3B, 3D, and 3E exhibit homology to phage sequences, suggesting a phage derivation for 4 or more kb of DNA in region 3. Homology to putative open reading frames in two M. leprae cosmids was also observed including homology to a putative bid gene encoding a protein involved in biotin synthesis. Also of interest was homology between ORF3A and an MTB sequence (mce) associated with cell invasion and intracellular survival.

Southern blot analysis with segments of region 3 deleted in BCG revealed that prototype lab strains of virulent MBV and MTB all carry deletion region 3 DNA. However, clinical isolates from PHRI are highly polymorphic or deleted in region 3.

This region contains many large direct and indirect repeats and, as mentioned above, at least 2 ORFs are homologous to phage sequences including homology to DNA invertases or recombinases. The repetitive nature of this region and the possible presence of a DNA recombinase could explain the polymorphisms observed in this region.

5

The sequence of a small region, estimated to be much less than 200 bp and located close to 9400 bp in Figure 3, was recalcitrant to automated sequencing and remains to be determined. Therefore, the base pair coordinates given in the region 3 map (Figure 6) 3' to the 9kb marker are approximations. The precise sequence determination of region is likely to effect the length of open reading frames 3H and 3L.

10

The foregoing subtractive analysis identified 3 regions in virulent M. bovis and M. tuberculosis prototype strains which are deleted in the avirulent BCG strain. The deletion located in region 2 may not have arisen in the original BCG Pasteur strain as this region is only deleted in strains derived from the original Pasteur strain after 1925. Region 3 is present in virulent MTB and MBV lab prototype strains (H37Rv, Erdman) and is highly polymorphic and at least partially deleted in the majority of MTB clinical isolates tested. Region 1 is apparently conserved and intact in all virulent MBV and MTB strains tested to date while all avirulent BCG strains tested to date are missing approximately 9kb from region 1.

## 20

15

## Example 2

Screening and Identification of an Avirulent Mycobacterium

The 32 D labeled subtraction probe obtained in Example 1, was

The <sup>32</sup> P labeled subtraction probe obtained in Example 1, was used to probe EcoRI and BamHI restricted chromosomal DNAs from BCG Connaught, *Mycobacterium bovis*, and various strains of *Mycobacterium tuberculosis* in a Southern blot. The hybridization was performed at 70°C in 6X SSC overnight.

25

The resulting Southern blot is illustrated in Figure 8. The probe showed no labeling of BCG reflecting the presence of all three deletions, while the other strains were labeled.

30

The above examples are provided to illustrate the invention but not to limit its scope. Other variants of the invention will be readily apparent to one of ordinary skill in the art and are encompassed by the appended claims. All publications, patents, and patent applications cited herein are hereby incorporated by reference.

WO 96/25519 PCT/US96/01938

## WHAT IS CLAIMED IS:

A marker for an avirulent mycobacterium, said marker comprising 1. a first nucleic acid that specifically hybridizes under stringent conditions with a second nucleic acid or a complement of said second nucleic acid where said second nucleic acid or complement of said second nucleic acid is selected from the group consisting of BCGala, BCGalb, BCGa2a, BCGa2b, BCGa3a, BCGa3b, BCGalab, BCGa2ab, BCGa3ab, BCGa1, BCGa2, and BCGa3.

The marker of claim 1, wherein said marker specifically hybridizes 2. under stringent conditions to a nucleic acid from BCG, but not to a nucleic acid from Mycobacterium tuberculosis or Mycobacterium bovis, or where said marker specifically hybridizes under stringent conditions to a nucleic acid from Mycobacterium tuberculosis or Mycobacterium bovis, but not to a nucleic acid from BCG.

1 2

3

4

5

1

2

3

4

1

2

1

2

3

4

5

6

7

8

9

10

11

- The marker of claim 2, wherein said marker comprises a 3. subsequence of a nucleic acid where said nucleic acid is selected from the group consisting of BCGala, BCGalb, BCGala, BCGalab, BCGalab, BCGalab, BCG△2ab, BCG△3ab, BCG△1, BCG△2, and BCG△3.
- 1 The marker of claim 2, wherein said marker is selected from the 4. group consisting of BCGala, BCGalb, BCGa2a, BCGa2b, BCGa3a, BCGa3b, 2 BCG△lab, BCG△2ab, BCG△3ab, BCG△1, BCG△2, and BCG△3. 3
  - The marker of claim 2, wherein said marker comprises a nucleic 5. acid having at least 90 percent sequence identity with a sequence selected from the group consisting of BCGala, BCGalb, BCGala, BCGalab, BCGalab, BCGalab, BCG△2ab, BCG△3ab, BCG△1, BCG△2, and BCG△3.
  - The marker of claim 2, wherein said marker comprises a 6. radioactive nucleotide probe.

1

4 5

1 The marker of claim 2, wherein said subsequence is a sequence 7. selected from an open reading frame of a deletion, said deletion being selected from the 2 group consisting of BCGa1, BCGa2, BCGa3. 3 1 A polypeptide encoded by a subsequence of a deletion sequence 8. selected from the group consisting of BCGa1, BCGa2, and BCGa3. 2 The polypeptide of claim 8, wherein the subsequence is selected 1 9. from an open reading frame (ORF) of a deletion, said deletion being selected from the 2 group consisting of BCGa1, BCGa2, BCGa3. 3 1 An antibody that binds specifically to the polypeptide of claim 8. 10. 1 A recombinant cell comprising a first nucleic acid that hybridizes 11. under stringent conditions with a second nucleic acid or a complement of said second 2 nucleic acid where said second nucleic acid or complement of said second nucleic acid is 3 selected from the group consisting of BCGala, BCGalb, BCGa2a, BCGa2b, BCGa3a, 4 BCGa3b, BCGa1ab, BCGa2ab, BCGa3ab, BCGa1, BCGa2, and BCGa3. 5 1 The recombinant cell of claim 11, wherein the cell is a 12. 2 Mycobacterium. 1 The cell of claim 11, wherein the cell expresses a polypeptide 13. encoded by an intact open reading frame from BCGa1, BCGa2, and BCGa3. 2 The cell of claim 11, wherein said cell is a mycobacterium having 14. one or more deletions in the genomic regions selected from the group consisting of 2 BCGa1, BCGa2, and BCGa3, wherein said deletions result in the attenuation of an 3 otherwise virulent strain of mycobacterium and wherein said deletions are present in up to two of said regions.

1	15. The mycobacterium of claim 14, wherein said deletions comprise a
2	deletion selected from the group consisting of BCGa1, BCGa2, and BCGa3.
1	16. A method of distinguishing between an attenuated and a virulent
2	mycobacterium, said method comprising detecting the presence or absence of a first
3	nucleic acid that hybridizes under stringent conditions with a second nucleic acid or a
4	complement of said second nucleic acid where said second nucleic acid or complement of
5	said second nucleic acid is selected from the group consisting of BCGala, BCGalb,
6	BCGa2a, BCGa2b, BCGa3a, BCGa3b, BCGa1ab, BCGa2ab, BCGa3ab, BCGa1,
7	BCG△2, and BCG△3.
1	17. The method of claim 16, wherein said first nucleic acid analisms.
2	To, wherein said this indicate acid specifically
3	hybridizes under stringent conditions to a nucleic acid from BCG, but not to a nucleic acid from Mycobacterium tuberculosis or Mycobacterium bovis, or where said first
4	nucleic acid specifically hybridizes under stringent conditions to a nucleic acid from
5	Mycobacterium tuberculosis or Mycobacterium bovis, but not to a nucleic acid from
6	BCG.
1	18. The method of claim 17 wherein said first sequence is a malified
2	18. The method of claim 17, wherein said first sequence is amplified prior to detection.
-	prior to detection.
1	19. The method of claim 17, wherein said first sequence is amplified
2	by the polymerase chain reaction.
1	20. A method of claim 17, wherein said detecting comprises a Southern
2	20. A method of claim 17, wherein said detecting comprises a Southern blot.
1	21. A method of claim 17, wherein said detecting comprises detecting a
2	polypeptide encoded by said first nucleic acid.

WO 96/25519

5

6

BCG∆2, and BCG∆3.

) PCT/US96/01938

	44
2	22. The method of claim 21, wherein the polypeptide is encoded by an
3	intact open reading frame of a nucleotide sequence selected from the group consisting of
4	BCG△1, BCG△2, and BCG△3.
1	23. The method of claim 21, wherein the polypeptide is visualized by
2	antibody hybridization.
1	24. A method for determining whether an attenuated or a virulent
2	Mycobacterium is present in a sample comprising:
3	providing a first nucleic acid that hybridizes under stringent conditions
4	with a second nucleic acid or a complement of said second nucleic acid where said
5	second nucleic acid or complement of said second nucleic acid is selected from the group
6	consisting of BCGala, BCGalb, BCGa2a, BCGa2b, BCGa3a, BCGa3b, BCGalab,
7	BCG\(\Delta\)2ab, BCG\(\Delta\)3ab, BCG\(\Delta\)1, BCG\(\Delta\)2, and BCG\(\Delta\)3; and
8	hybridizing said first nucleic acid to the biological sample.
1	25. The method of claim 24, wherein said first nucleic acid specifically
2	hybridizes under stringent conditions to a nucleic acid from BCG, but not to a nucleic
3	acid from Mycobacterium tuberculosis or Mycobacterium bovis, or where said first
4	nucleic acid specifically hybridizes under stringent conditions to a nucleic acid from
5	Mycobacterium tuberculosis or Mycobacterium bovis, but not to a nucleic acid from
6	BCG.
1	26. A method of producing an attenuated Mycobacterium species, said
2	method comprising deleting from the genomic DNA of a virulent mycobacterium a first
3	nucleic acid that specifically hybridizes under stringent conditions with a second nucleic
4	acid or a complement of said second nucleic acid where said second nucleic acid or
_	and the second second of

complement of said second nucleic acid is selected from the group consisting of BCGa1,

WO 96/25519 PCT/US96/01938

1 / 63

099

GCCATACATG

720

CAGGCGTCGA AACCAATGGT GAAGCCGGAG GGCCCGGTCA ACAGACCACT

GGTCAAGTTC AGCACCTTCT ACATCAGTGG

CGGGTCGGCA TGGAAGACCC

GTAATCGGCA

780

TTCTCGAGGA GGCGCCGACA

GCGGCACCGG

CAGGTTCACC

CGCGCATTCA

AGACAAGCCG

5505500500

22 09 180 360 420 120 240 300 480 540 009 of Page GAATTCCTGC GCACCCTGAT CCTGTCGCTG GTGGCAATGA CTCATCCAGA TCAGGTGAAT GCTTCCGCAC GATGGGCGAG GATGAAAGTC SECECECCE GAGCCCTGTC CEGCETEECC GAATACGAGA AGTACCECGA ACECEGTECC SACCTACCCC CGCTGCCAAC GCTTTTCGTC GTCGTCGACG AGTTCGCCGA GCTGTTGCAG AGTCACCCGG ACTTCATCGG GCTGTTCGAC CGGATCTGCC GCGTCGGGCG GTCGCTGAGG GICCAICIGC IGCIGGCIAC CCAGICGCIG CAGACCGGCG GIGIICGCAI CGACAAACIG CGCACCACCA GCTCTCATGA ATCCAAGGCG ACCAACAAGG AGAGCGGTGT CGGGTTTCTC TCGTCAGCCG GAATGGAAAA GACAGGCCGG CCGACTTCAA AGGTGGTTCA ACCTTCCTGG TCGTCACCAA CATGGCCGAG GAAGCCGAGC TCGATCCTCC FIGURE 1-1 GTGTTGACCG GAGAACTCGA TCGGCGCCAG GAGCCAAACC TGACATATCG AATCGCATTG CACCGGAGGC GCAGTACATC

ACTGCCGCTG

CTCCTGCTCA

						2/6	3						
2 of 22	840	006	096	1020	1080	1140	1200	1260	1320	1380	1440	1500	1560
Page	GGAGGAGCGG CGCCAACGGC	CGGCGCGAT GACTGCTGAA	CCGGAAGTAC GGACGCTGCG CGAGGTTGTG CTGGACCAGC TCGGCACTGC TGAATCGCGT	CGCTCAACGA GCTCATCGCC	TGGATGAACC GCGCCGCCAT	GCAACATCGG TATTGGGGGC	TGATGTCGGC CGCCGCCACA	GTGGCGCGG GCTGATCTAT	CCGAGCCCGA CAAGGTCAAC	AAACCACCTT CAAGGAACAC	CTGCGTGACG ATCCAAGTCA ACCCGTTGCG	GGCCCGGTTT TGTCGGCGAG	TICAAGAT CIGGCCGCCC AGGGGCTGGG GITCGGCGTC
FIGURE 1-2	GATGCAAAGC GCAGCGATGA	AGCGCAGCGA TGAGGAGGAG	CGAGGTTGTG CTGGACCAGC	GCCGTTGACC AATCCGGTCC	GCGATTTGCC CTGGGGATCA	AGACGTTTCC GGGGCCGGCG	GACGCTACTG CAGACGATGG	GTTCTATTGC ATCGACCTAG	CGGTGGGGTA GCCAATCGGT	AGCCGTCATG CGGCAACGGG	ACCGGCAG	CTTTCTGATC ATCGACGGAT	
	CCGTGACCCG CGCCGGCGAC GATGCAAAGC GCAGCGATGA GGAGGAGCGG	CCGCCCCGC GACGATGCAA AGCGCAGCGA	CCGGAAGTAC GGACGCTGCG	GCGTACAAGA TGTGGCTGCC GCCGTTGACC AATCCGGTCC CGCTCAACGA GCTCATCGCC	CGTGATCGGC GACAACCCCT GCGATTTGCC	CTACAGGATG TGTGGGGCGT AGACGTTTCC GGGGCCGGCG GCAACATCGG	GCACCTCAAA CCGGGAAGTC GACGCTACTG CAGACGATGG TGATGTCGGC CGCCGCCACA	CACTCACCGC GCAACGTTCA GTTCTATTGC	CTCGAAAACC TTCCACACGT CGGTGGGGTA	CGGGTGGTCG CAGAGATGCA AGCCGTCATG CGGCAACGGG AAACCACCTT CAAGGAACAC	CGAGTGGGCT CGATCGGGAT GT	TCCGATCCAT ACGGCGACGT CTTTCTGATC	TTCCCCGACC TTGAGGGGCA GG

FIGURE 1-3 Page	3 of 22
CACGICATCA TCTCCACGCC ACGCTGGACA GAGCTGAAGT CGCGTGTTCG CGACTACCTC	1620
GGCACCAAGA TCGAGTTCCG GCTTGGTGAC GTCAATGAAA CCCAGATCGA CCGGATTACC	1680
CGCGAGATCC CGGCGAATCG TCCGGGTCGG GCAGTGTCGA TGGAAAAGCA CCATCTGATG	1740
ATCGGCGTGC CCAGGTTCGA CGGCGTGCAC AGCGCCGATA ACCTGGTGGA GGCGATCACC	1800
GCGGGGGTGA CGCAGATCGC TTCCCAGCAC ACCGAACAGG CACCTCCGGT GCGGGTCCTG	1860
CCGGAGCGTA TCCACCTGCA CGAACTCGAC CCGAACCCGC CGGGACCAGA GTCCGACTAC	1920
CGCACTCGCT GGGAGATTCC GATCGGCTTG CGCGAGACGG ACCTGACGCC GGCTCACTGC	1980
CACATGCACA CGAACCCGCA CCTACTGATC TTCGGTGCGG CCAAATCGGC CAAGACGACC	2040
ATTGCCCACG CGATCGCGCG CGCCATTTGT GCCCGAAACA GTCCCCAGCA GGTGCGGTTC	2100
ATGCTCGCGG ACTACCGCTC GGGCCTGCTG GACGCGGTGC CGGACACCCCA TCTGCTGGGC	2160
GCCGGCGCGA TCAACCGCAA CAGCGCGTCG CTAGACGAGG CCGCTCAAGC ACTGGCGGTC	2220
AACCTGAAGA AGCGGTTGCC GCCGACCGAC CTGACGACGG CGCAGCTACG CTCGCGTTCG	2280
TGGTGGAGCG GATTTGACGT CGTGCTTCTG GTCGACGATT GGCACATGCA GCCGTGGGTG	2340

						••							
1 of 22	2400	2460	2520	2580	2640	2700	2760	2820	2880	2940	3000	3060	3120
Page 4	ATTGCCGGCG GCGGCAGATA	ACCTGTCAGA TGAGCCAGGC TTACAAGGCA ACCATGGACA	GTTCCTTTCG GGCGAGAAGC	CCCTGGCCAG GCATTTCTCG	CGAGCCTCCA GAAGAAGTGT	CGGTGTAGCA GGACCCGAGC	TTTGTTTCCG GCTATAACCG	GAGGGAAGAA GTAGGCAAAT GGAAAAAATG	GCGACAACGC TCTGCACGGC	TGGTTCCCGC GGGGGCCGAT	GCATCCAATT GCTGGCTTCC	CGGTCCAGGA CGTCGCCCGC	CCTAATAGGC CCCCAACACA
FIGURE 1-4	CCGCCGGGGG GATGCCGCCG ATGGCACCGC TGGCCCCGTT	TCGGGTTGCA CATCATTGTC ACCTGTCAGA TGAGCCAGGC	AGTTCGTCGG CGCCGCATTC GGGTCGGGCG CTCCGACAAT GTTCCTTTCG GGCGAGAAGC	AGGAATTCCC ATCCAGTGAG TTCAAGGTCA AGCGGCGCCC	TCTCGCCAGA CGGCAAAGAG GTCATCCAGG CCCCCTACAT	TCGCAGCACC CCCAAGCGCC GGTTAAGATT ATTTCATTGC CGGTGTAGCA GGACCCGAGC	TCAGCCCGGT AATCGAGTTC GGGCAATGCT GACCATCGGG	AACGGTTTGT GTACGGGATA CAAATACAGG GAGGGAAGAA	TCACATGATC CGATGCTTGC CGACATTGGC ACGCAAGTGA GCGACAACGC	GTGACGGCCG GCTCGACGGC GCTGACGTCG GTGACCGGGC TGGTTCCCGG GGGGGCCGAT	GAGGICTCCG CCCAAGCGGC GACGGCGTTC ACATCGGAGG GCATCCAATT	AATGCATCGG CCCAAGACCA GCTCCACCGT GCGGGCGAAG	ACCTATICGC AAATCGACGA CGGCGCCGCC GGCGTCTTCG CCTAATAGGC CCCCAACACA

		FIC	FIGURE 1-5		Page 5	of 22	
TCGGAGGGAG	TCGGAGGGAG TGATCACCAT GCTGTGGCAC GCAATGCCAC CGGAGCTAAA TACCGCACGG	GCTGTGGCAC	GCAATGCCAC	CGGAGCTAAA	TACCGCACGG	3180	
CTGATGGCCG	CTGATGGCCG GCGCGGGTCC GG	GGCTCCAATG	CTTGCGGCGG	CCGCGGGATG	GCAGACGCTT	3240	
TCGGCGGCTC	TCGGCGGCTC TGGACGCTCA GG	GGCCGTCGAG	TTGACCGCGC	CCGTCGAG TTGACCGCGC GCCTGAACTC TCTGGGAGAA	TCTGGGAGAA	3300	
GCCTGGACTG	GCCTGGACTG GAGGTGGCAG CGACAAGGCG CTTGCGGCTG CAACGCCGAT GGTGGTCTGG	CGACAAGGCG	CTTGCGGCTG	CAACGCCGAT	GGTGGTCTGG	3360	
CTACAAACCG	CTACAAACCG CGTCAACACA GG	GGCCAAGACC	CGTGCGATGC AGGCGACGGC		GCAAGCCGCG	3420	
GCATACACCC	GCATACACCC AGGCCATGGC CA	CACGACGCCG	TCGCTGCCGG	AGATCGCCGC CAACCACATC	CAACCACATC	3480	5/6
ACCCAGGCCG	ACCCAGGCCG TCCTTACGGC CACCAACTTC	CACCAACTTC	TTCGGTATCA	TTCGGTATCA ACACGATCCC GATCGCGTTG	GATCGCGTTG	3540	3
ACCGAGATGG	ACCGAGATGG ATTATTTCAT CCGTATGTGG		AACCAGGCAG	AACCAGGCAG CCCTGGCAAT GGAGGTCTAC	GGAGGTCTAC	3600	
CAGGCCGAGA	CAGGCCGAGA CCGCGGTTAA CACGCTTTTC		GAGAAGCTCG	AGCCGATGGC	GTCGATCCTT	3660	
GATCCCGGCG	GATCCCGGCG CGAGCCAGAG CA	CACGACGAAC	CCGATCTTCG	CGACGAAC CCGATCTTCG GAATGCCCTC CCCTGGCAGC	CCCTGGCAGC	3720	
TCAACACCGG	TCAACACCGG TTGGCCAGTT GCCGCCGGCG GCTACCCAGA CCCTCGGCCA ACTGGGTGAG	ອວອອວວອວວອ	GCTACCCAGA	CCCTCGGCCA	ACTGGGTGAG	3780	
ATGAGCGGCC	ATGAGCGGCC CGATGCAGCA GCTGACCCAG	GCTGACCCAG	CCGCTGCAGC AGGTGACGTC		GTTGTTCAGC	3840	
CAGGTGGGCG	CAGGTGGGCG GCACCGGCGG CGGCAACCCA		GCCGACGAGG	AAGCCGCGCA	GATGGGCCTG	3900	

						6/6	3						
of 22	3960	4020	4080	4140	4200	4260	4320	4380	4440	4500	4560	4620	4680
Page 6	CAGCGCGGC	CCGCACGCCG	CGGCTGCTGC	GCCAGGGTGC	CGCAGGAGCG	CCGTAATGAC	GGAAGGTAAA	CGCGCAGGAG	GGTGGAGTCG	CCAGGCCGCG	GATCTCGACG	GCAGGCGCTG	AACATGACAG
	GATCAGGCCC CAGCGCGGGC	GGTCGTTGAC	CTGATGTCTC AGCTGATCGA AAAGCCGGTT TGCCCCCTCG GTGATGCCGG CGGCTGCTGC	CGGATCGTCG GCGACGGGTG GCGCCGCTCC GGTGGGTGCG GGAGCGATGG GCCAGGGTGC	GGTCGCGCCG GCACCGCTCG	TGGTGAGCTC	AACAGACTTC CCGGCCACCC GGGCCGGAAG ACTTGCCAAC ATTTTGGCGA GGAAGGTAAA	GAGAGAAAGT AGTCCAGCAT GGCAGAGATG AAGACCGATG CCGCTACCCT CGCGCAGGAG	AGATCGACCA GGTGGAGTCG	GGACGGCCGC	GTGGTGCGCT TCCAAGAAGC AGCCAATAAG CAGAAGCAGG AACTCGACGA GATCTCGACG	AATACTCG AGGGCCGACG AGGAGCAGCA GCAGGCGCTG	ACGGAGCAAA AACATGACAG
FIGURE 1-6	CTGGCTGGTG	GGCGCAGGTG	тессссстсв	GGTGGGTGCG		AGAGGACGAC	ACTTGCCAAC	AAGACCGATG	CTGAAAACCC	9929929299	CAGAAGCAGG	AGGGCCGACG	CCGCTAAT ACGAAAAGAA
	GAACCATCCG	GTCGCTACCT	AAAGCCGGTT	CCCCCCTCC	GGCCGGGTCT	ACTGGGACGA	GGGCCGGAAG	GGCAGAGATG	CTCCGGCGAC	CCAGTGGCGC	AGCCAATAAG		ACCCGCTAAT
	CTCGGCACCA GTCCGCTGTC GAACCATCCG CTGGCTGGTG	GCGGGCCTGC TGCGCGCGA GTCGCTACCT	AGCTGATCGA	GCGACGGGTG	GCAATCCGGC GGCTCCACCA GGCCGGGTCT	TGAAGAAGAC GACGAGGACG ACTGGGACGA	CCGGCCACCC	AGTCCAGCAT	GCAGGTAATT TCGAGCGGAT CTCCGGCGAC	CGTTGCAGGG	TCCAAGAAGC	AATATTCGTC AGGCCGGCGT CC	TCCTCGCAAA TGGGCTTCTG AC
	CTCGGCACCA	GCGGGCCTGC	CTGATGTCTC	CGGATCGTCG	GCAATCCGGC	TGAAGAAGAC	AACAGACTTC	GAGAGAAAGT	GCAGGTAATT	ACGGCAGGTT	GTGGTGCGCT	AATATTCGTC	TCCTCGCAAA

		FIC	FIGURE 1-7		Page 7	of 22	
GCAGTG	GAATTTCGCG	AGCAGCAGTG GAATTTCGCG GGTATCGAGG	CCGCGGCAAG	CCGCGGCAAG CGCAATCCAG GGAAATGTCA	GGAAATGTCA	4740	
CATTCA	CGTCCATTCA TTCCCTCCTT	GA	CGAGGGA AGCAGTCCCT	GACCAAGCTC GCAGCGGCCT	GCAGCGGCCT	4800	
CGGTAG	CGGTTCGGAG	GGGGCGGTAG CGGTTCGGAG GCGTACCAGG GTGTCCAGCA AAAATGGGAC GCCACGGCTA	GTGTCCAGCA	AAAATGGGAC	GCCACGGCTA	4860	
GCTGAA	CAACGCGCTG	CCGAGCTGAA CAACGCGCTG CAGAACCTGG CGCGGACGAT	CGCGGACGAT	CAGCGAAGCC GGTCAGGCAA	GGTCAGGCAA	4920	
TTCGAC	TGGCTTCGAC CGAAGGCAAC GT	GTCACTGGGA	TGTTCGCATA	TGTTCGCATA GGGCAACGCC	GAGTTCGCGT	4980	
TAGCGA	AGAATAGCGA AACACGGGAT	CGGGCGAGTT	CGACCTTCCG	CGACCTTCCG TCGGTCTCGC CCTTTCTCGT	CCTTTCTCGT	5040	7/ (
ATACGT	TTGAGCGCAC	GTTTATACGT TTGAGCGCAC TCTGAGAGGT TGTCATGGCG GCCGACTACG ACAAGCTCTT	TGTCATGGCG	GCCGACTACG	ACAAGCTCTT	5100	63
CCGCAC	GAAGGTATGG	CCGGCCGCAC GAAGGTATGG AAGCTCCGGA	CGATATGGCA GCGCAGCCGT		TCTTCGACCC	5160	
GCTTCG	CAGTGCTTCG TTTCCGCCGG CGCCCGCATC		GGCAAACCTA	CCGAAGCCCA	ACGGCCAGAC	5220	
ממממפ	ACGTCCGACG	TCCGCCCCCG ACGTCCGACG ACCTGTCGGA GCGGTTCGTG TCGGCCCCCGC CGCCGCCACC	GCGGTTCGTG	TCGGCCCCGC	CGCCGCCACC	5280	
CCCCCA	CCTCCGCCTC	CCCACCCCCA CCTCCGCCTC CGCCAACTCC GATGCCGATC GCCGCAGGAG AGCCGCCCTC	GATGCCGATC	GCCGCAGGAG	AGCCGCCCTC	5340	
GAACCG	GCCGCATCTA	GCCGGAACCG GCCGCATCTA AACCACCCAC	ACCCCCCATG	CCCATCGCCG	GACCCGAACC	5400	
CCACCC	GGCCCCACCC AAACCACCCA CA	CACCCCCCAT	GCCCATCGCC	CCCCCCAT GCCCATCGCC GGACCCGAAC CGGCCCCACC	CGGCCCCACC	5460	

8	/	63
---	---	----

	FIC	FIGURE 1-8		∞	of 22
CAAACCACCC ACACCTCCGA	TGCCCATCGC	CGGACCTGCA	TGCCCATCGC CGGACCTGCA CCCACCCCAA CCGAATCCCA	CCGAATCCCA	5520
GITGGCGCCC CCCAGACCAC	CGACACCACA	CGACACCACA AACGCCAACC	GGAGCGCCGC	AGCAACCGGA	5580
ATCACCGGCG CCCCACGTAC	CCTCGCACGG	GCCACATCAA	CCTCGCACGG GCCACATCAA CCCCGGCGCA CCGCACCAGC	CCGCACCAGC	5640
ACCGCCCTGG GCAAAGATGC	CAATCGGCGA	ACCCCCGCCC	CAATCGGCGA ACCCCGCCC GCTCCGTCCA GACCGTCTGC	GACCGTCTGC	5700
GTCCCCGGCC GAACCACCGA	CCCGGCCTGC	CCCGGCCTGC CCCCCAACAC	TCCCGACGTG	ອອອອລລອລອລ	5760
TCACCGCTAT CGCACAGACA	CCGAACGAAA	CGTCGGGAAG	GTAGCAACTG	GTCCATCCAT	5820
CCAGGCGCGG CTGCGGGCAG	AGGAAGCATC	CGGCGCGCAG	AGGAAGCATC CGGCGCGCAG CTCGCCCCCG GAACGGAGCC	GAACGGAGCC	5880
CTCGCCAGCG CCGTTGGGCC	AACCGAGATC	GTATCTGGCT	AACCGAGATC GTATCTGGCT CCGCCCACCC GCCCCGCGCC	ವದ್ದದ್ದದ್ದದ್ದ	5940
GACAGAACCT CCCCCCAGCC	CCTCGCCGCA	CCTCGCCGCA GCGCAACTCC GGTCGGCGTG		ccgagcgacg	0009
CGTCCGACCC CGATTTAGCC	GCCCAACATG	CCGCGGCGCA	GCCCAACATG CCGCGCGCA ACCTGATTCA ATTACGGCCG	ATTACGGCCG	0909
CAACCCACTG GCGGTCGTCG	CCGCAAGCGT	GCAGCGCCGG	CCGCAAGCGT GCAGCGCCGG GATGCTCGAC GCGACACAAG	GCGACACAAG	6120
AAATCCTTAA GGCCGGCGGC	CAAGGGGCCG	AAGGTGAAGA	CAAGGGGCCG AAGGTGAAGA AGGTGAAGCC CCAGAAACCG	CCAGAAACCG	6180
AAGGCCACGA AGCCGCCCAA	AGTGGTGTCG	CAGCGCGGCT	GGCGACATTG	GGTGCATGCG	6240

						9/6	63						
9 of 22	6300	6360	6420	6480	6540	0099	0999	6720	6780	6840	0069	0969	7020
Page 9	CCTGCACGCT	CAAAGGTGGG	GCGGCCCGAC	GGTAGGGCGA	CTACAACGAC	ACCGGAATAC	TCCTGCGTCG	GCTGACCCGC	CGACGGCGCA	TTTGGCGAGC	AGTTAAAGAC	GCCGTGGGAC	CTACAAGCGC
	TTGACGCGAA TCAACCTGGG CCTGTCACCC GACGAGAAGT ACGAGCTGGA CCTGCACGCT	TCGTCGGTCT	GCTGGCAAAA CCACGCTGAC AGCAGCGTTG GGGTCGACGT TGGCTCAGGT GCGGGCCGAC	CGGATCCTGG CTCTAGACGC GGATCCAGGC GCCGGAAACC TCGCCGATCG GGTAGGGCGA	AGCTGTCGCA CTACAACGAC	TGCTGCCGGC	AGCTCGGCGC AGCGCGCGCT CAGCGACGCC GACTGGCATT TCATCGCCGA TCCTGCGTCG	AGGTTTTACA ACCTCGTCTT GGCTGATTGT GGGGCCGGCT TCTTCGACCC GCTGACCCGC	GTCGTGGCAA GTGTCTCAAT CGACGGCGCA	CGCAACAACG GTTACCAAGA	CCGGGAGAAC CCAATGTCGC AGTTAAAGAC	CTGGTGCGGC ATTTCGAACA GCAAGTTCAA CCCGGCCGGG TCGTGGTCAT GCCGTGGGAC	TCGACCCTAT
FIGURE 1-9	GACGAGAAGT	CAGATCGCCG	GGGTCGACGT	GCCGGAAACC	GCAGAAAAAG	AATCTGGAAG	GACTGGCATT	GGGGCCGGCT	GTCGTGGCAA	CGCAACAACG	CCGGGAGAAC	ອອອວວອອວວວ	CTCGACTTGC
FIC	CCTGTCACCC	CGGGTCGTAT	AGCAGCGTTG	GGATCCAGGC	TGATGTGCTT	CAATGCGGTC	CAGCGACGCC	GGCTGATTGT	CGGTGTCGTG	GGACTGGTTG	TCACATCATG	GCAAGTTCAA	CGAGATTTCA
	TCAACCTGGG	CGAGTCCGCC GCAATCCCCG CGGGTCGTAT	CCACGCTGAC	CTCTAGACGC	CAATCGGGCG CGACCATCGC TGATGTGCTT	ATCCGCGCAC ACACTAGCGT CAATGCGGTC	AGCGCGCGCT	ACCTCGTCTT	GGCGTGCTGT CCACGGTGTC CGGTGTCGTG	CAACAGGCGT CGGTCGCGTT GGACTGGTTG	CGCGCATGCG TGGTCATCAA TCACATCATG	ATTTCGAACA	AGGCACATTG CGGCCGGAAC CGAGATTTCA
	TTGACGCGAA	CGAGTCCGCC	GCTGGCAAAA	CGGATCCTGG	CAATCGGGCG	ATCCGCGCAC	AGCTCGGCGC	AGGTTTTACA	GGCGTGCTGT	CAACAGGCGT	CGCGCATGCG	CTGGTGCGGC	AGGCACATTG

FIGURE 1-10	Page 10	of 22
AAGGTCCTCG AATTGGCCGC AGCGCTATCC GACGATTTCG AGAGGGCTGG ACGTCGTTGA	GG ACGTCGTTGA	7080
GCGCACCTGC TGTTGCTGCT GGTCCTACCG CCGCGGGGC AACCGCTGCG	CG CGCCTGCCA	7140
CCACCCGGGT GACGATCCTG ACCGGCAGAC GGATGACCGA TTTGGTACTG CCAGCGGCGG	TG CCAGCGGCGG	7200
TGCCGATGGA AACTTATATT GACGACACCG TCGCGGTGCT TTCCGAGGTG TTGGAAGACA	TG TTGGAAGACA	7260
CGCCGGCTGA TGTACTCGGC GGCTTCGACT TTACCGCGCA AGGCGTGTGG	GG GCGTTCGCTC	7320
GTCCCGGATC GCCGCCGCTG AAGCTCGACC AGTCACTCGA TGACGCCGGG	GG GTGGTCGACG	7380
GGTCACTGCT GACTCTGGTG TCAGTCAGTC GCACCGAGCG CTACCGACCG TTGGTCGAGG	CG TTGGTCGAGG	7440
ATGTCATCGA CGCGATCGCC GTGCTTGACG AGTCACCTGA GTTCGACCGC ACGGCATTGA	GC ACGCCATTGA	7500
ATCGCTTTGT GGGGGGGGG ATCCCGCTTT TGACCGCGCC CGTCATCGGG	GG ATGGCGATGC	7560
GGGCGTGGTG GGAAACTGGG CGTAGCTTGT GGTGGCCGTT GGCGATTGGC ATCCTGGGGA	GC ATCCTGGGGA	7620
TCGCTGTGCT GGTAGGCAGC TTCGTCGCGA ACAGGTTCTA CCAGAGCGGC CACCTGGCCG	GC CACCTGGCCG	1680
AGTGCCTACT GGTCACGACG TATCTGCTGA TCGCAACCGC CGCAGCGCTG	TG GCCGTGCCGT	7740
TGCCGCGCGG GGTCAACTCG TTGGGGGCGC CACAAGTTGC CGGCGCCGCT ACGGCCGTGC	CT ACGCCGTGC	7800

		F	FIGURE 1-11		Page 1	11 of 22	
TGTTTTTGAC	TGTTTTTGAC CTTGATGACG CGGGGGGGCC	റാളൊള്ളൊ	CTCGGAAGCG	TCATGAGTTG	GCGTCGTTTG	7860	
CCGTGATCAC	CCGTGATCAC CGCTATCGCG GTCATCGCGG	GTCATCGCGG	CCGCCGCTGC	CTTCGGCTAT GGATACCAGG	GGATACCAGG	7920	
ACTGGGTCCC	5555555555	ATCGCATTCG	GGCTGTTCAT	ACTGGGTCCC CGCGGGGGG ATCGCATTCG GGCTGTTCAT TGTGACGAAT GCGGCCAAGC	GCGCCCAAGC	7980	
TGACCGTCGC	GGTCGCGCGG	ATCGCGCTGC	CGCCGATTCC	TGACCGTCGC GGTCGCGCGG ATCGCGCTGC CGCCGATTCC GGTACCCGGC GAAACCGTGG	GAAACCGTGG	8040	
ACAACGAGGA	ACAACGAGGA GTTGCTCGAT CCCGTCGCGA	CCCGTCGCGA	CCCCGGAGGC	TACCAGCGAA GAAACCCCGA	GAAACCCCGA	8100	
CCTGGCAGGC	CCTGGCAGGC CATCATCGCG TCGGTGCCCG	TCGGTGCCCG		CGTCCGCGGT CCGGCTCACC GAGCGCAGCA	GAGCGCAGCA	8160	11/6
AACTGGCCAA	GCAACTTCTC	ATCGGATACG	TCACGTCGGG	AACTGGCCAA GCAACTTCTC ATCGGATACG TCACGTCGGG CACCCTGATT CTGGCTGCCG	CTGGCTGCCG	8220	3
GTGCCATCGC	GTGCCATCGC GGTCGTGGTG CGCGGCACT	CGCGGGCACT	TCTTTGTACA	TCTTTGTACA CAGCCTGGTG	GTCGCGGGTT	8280	
TGATCACGAC	TGATCACGAC CGTCTGCGGA TTTCGCTCGC	TTTCGCTCGC	GGCTTTACGC	CGAGCGCTGG	TGTGCGTGGG	8340	
CGTTGCTGGC	GGCGACGGTC	GCGATTCCGA	CGGGTCTGAC	CGTTGCTGGC GGCGACGGTC GCGATTCCGA CGGGTCTGAC GGCCAAACTC ATCATCTGGT	ATCATCTGGT	8400	
ACCCGCACTA	ACCCGCACTA TGCCTGGCTG TTGTTGAGCG	TTGTTGAGCG	TCTACCTCAC GGTAGCCCTG	GGTAGCCCTG	GTTGCGCTCG	8460	
TGGTGGTCGG	TGGTGGTCGG GTCGATGGCT CACGTCCGGC	CACGTCCGGC	GCGTTTCACC	GGTCGTAAAA	CGAACTCTGG	8520	
AATTGATCGA	CGGCGCCATG	ATCGCTGCCA	TCATTCCCAT	AATTGATCGA CGGCGCCATG ATCGCTGCCA TCATTCCCAT GCTGCTGTGG ATCACCGGGG	ATCACCGGGG	8580	

						12/	63						
of 22	8640	8700	8760	8820	8880	8940	0006	0906	9120	9180	9240	9300	9360
12													
Page	GGCTGATTGG CGGTTCCTGA	CGCCCGACAA ATTGCTGCGA	TTGAGCGTGT GGCGCGTCCG GTAAATTTG CTCGATGGGG AACACGTATA GGAGATCCGG	CGATCCCA CCGGCTTGAG CGCAGCGGCC GCGAAATTGG	CAGCGGAACG GATTCGGTGG	TCGAATCGCT GGTCAGTGAC GGGCTGCCCG	GCGTGAAAGC CGCCCTGACT CGAACAGCAT CCAACATGAA CGCGGCGGCG GACGTCTATG	TGAGCCAGTA TGCATTCGGC TCGTCGGGCG	TCAGGCTACC CAGCTGCTGA	GGCCGCTGAG CTGGCACCCC	GIGITGITGC GACGGIGCCG CAACTCGTIC AGCIGGCICC GCACGCCGIT CAGAIGICGC	AAAACGCATC CCCCATCGCT CAGACGATCA GTCAAACCGC CCAACAGGCC GCCCAGAGCG	CGCTGAAAAA CCGGCCACCG
			AA	CG	CA	GG	CG	TG(		99	GC7	CCI	
FIGURE 1-12	CCGGTTCT GAGCCGGATC	ATACCTTCGG	CTCGATGGGG	CCGGCTTGAG	CGATCGCGGT		CCAACATGAA		GTCAGCCAAG	TCGGCGAGAC	AGCTGGCTCC	GTCAAACCGC	AGCTTGCCAG
	ATCCGGTTCT	GCAGGTTTGC	GTAAAATTTG	GTCGATCCCA	ccrcceecec	ATGCCAAGCA	CGAACAGCAT	GGAACCAGTT	TCGGTCGGTG	ACGACCCAGC	CAACTCGTTC	CAGACGATCA	ATGCCCGCAC
	TGTACGACAC GGTCCGCAAT AT	CAGAACATCG AGGACACGGC GC	GGCGCGTCCG	CAATGGCTGA ACCGTTGGCC GT	CCGGCCTCGT TTTTCCGCAG CCTCCGGCGC	TAGCAGCAAT CAACAAGACC ATGCCAAGCA	CGCCCTGACT	CGAAGACCGA TCAGTCACTG GGAACCAGTT	AAGGCCTGGC TGGCGTCGCC TCGGTCGGTG	GCACACCCGT GTCACAGGTC ACGACCCAGC	GACGGTGCCG	CCCCATCGCT	CGCAGGGCGG CAGCGGCCCA ATGCCCGCAC
	TGTACGACAC	CAGAACATCG	TTGAGCGTGT	CAATGGCTGA	CCGGCCTCGT	TAGCAGCAAT	GCGTGAAAGC	CGAAGACCGA	AAGGCCTGGC	GCACACCCGT	GTGTTGTTGC	AAAACGCATC	CGCAGGGCGG

						13/6	3						
of 22	9420	9480	9540	0096	0996	9720	9780	9840	0066	0966	10020	10080	10140
FIGURE 1-13 Page 13	AGCAAGCGGA GCCGGTCCAC GAAGTGACAA ACGACGATCA GGGCGACCAG GGCGACGTGC	AGCCGGCCGA GGTCGTTGCC GCGGCACGTG ACGAAGGCGC CGGCGCATCA CCGGGCCAGC	AGCCCGGCGG AGGCGTTCCC GCGCAAGCCA TGGATACCGG AGCCGGTGCC CGCCCAGCGG	CGAGTCCGCT GGCGGCCCCC GTCGATCCGT CGACTCCGGC ACCCTCAACA ACCACAACGT	TGTAGACCGG GCCTGCCAGC GGCTCCGTCT CGCACGCAGC GCCTGTTGCT GTCCTGGCCT	CGTCAGGATG CGGCGGCCAG GGCCCGGTCG AGCAACCCGG TGACGTATTG CCAGTACAGC	CAGTCCGCGA CGGCCACACG CTGGACGGCC GCGTCAGTCG CAGTGTGCGC TTGGTGCAGG	GCAATCTCCT GTGAGTGGGC AGCGTAGGCC CGGAACGCCC GCAGATGAGC GGCCTCGCGG	CCGGTAGCGG TGCTGGTCAT GGGCTTCATC AGCTCGAACC ACAGCATGTG CCGCTCATCG	CCCGGTGGAT TGACATCCAC CGGCGCCGGC GGCAACAAGT CGAGCAAACG CTGATCGGTA	GTGTCGGCCA GCTGAGCCGC CGCCGAGGGG TCGACGACCT CCAGCCGCGA CCGGCCCGTC	ATTTTGCCGC TCTCCGGAAT GTCATCTGGC TCCAGCACAA TCTTGGCCAC ACCGGGATCC	GAACTGGCCA ACTGCTCCGC GGTACCGATC ACCGCCCGCA GCGTCATGTC GTGGAAAGCC
	AGCA	AGCCC	AGCCC	CGAGI	TGTAG	CGICA	CAGTC	GCAAT	CCGGI	ອຄລລລ	GTGTC	ATTT	GAACT

						14/63	3						
14 of 22	10200	10260	10320	10380	10440	10500	10560	10620	10680	10740	10800	10860	10920
FIGURE 1-14 Page	GCCCAGGCTT GCACGGCCAA AACCGGGTAG GTGGCACAGC GTGCAATTTC GTCAACCGGG	ATTGCGTGAT CCGCGCTGGC CAAGTACACC TTATTCGGCA ATTCCATCCC GTCGGGTATG	TAGGCCAGCC CATAGCTGTT GGCCACGACG ATGGAACCGT CGGTGGTCAC CGCGGTGATC	CAGAAGAACC CGTAGTCGCC CGCGTTGTTG TCGGACGCGT TGAGCGCCCGC CGCGATGCGT	CGCGCCAACC GCAGCGCATC ACCGCGGCCA CGCTGGCGGG CGCTGGCAGC TGCAGTGGCG	GCGTCGCGTG CCGCCCGAGC CGCCGACACC GGGATCATCG ACACCGGCGT ACCGTCATCT	GCAGACTCGC TGCGATCGGG TTTGTCGATG TGATCGGTCG ACGGAGGCG GGCAGGAGGT	GCCGTCCGCG CCGAGGCCGC CCGCGTGCTC GGTGCCGCCG CCTTGTCCGA GGTAGCCACC	TGCGTCCGCC CAGTGGCAGT ATGCGGACCC CGGAAAAAA AAACTCGAGT GCGTTCTTCG	GAGGTTTCCA ATTCTTGGAT TCCAGCACCC GCTCAGCGGT CTCGGCGACC AGACTGACAT	TGGCCCCATG CGTCGCCGTG ACCAATGAAT TGATGGCGGT ATGGCGCTCA TCAGCATCCA	GGCTAGAGTC ATTCTCCAGG ATATCGATCT CCCGTTGAGC GCCATCCACA TTATTGCCGA	TATCGGATTT AGCTTGCTCA ATCAACCCGG CAATATGCCT GTGCCAGGTA ATCACCGTGG

		FIC	FIGURE 1-15		Page 15 of	.5 of 22	
AGATAATC	CGAGATAATC CTGCAGCGTC ATCAATTGAT	ATCAATTGAT	TGATGTTTGC	TGATGTTTGC ACCCAGGGCG	CCGTTGGCAG	10980	
TTGGCGGC	CATTGGCGGC GCCGCCGGAC CATAGGCCGC	CATAGGCCGC	CTTCGAAGAC	GTGGCCTTTC	TGCTGGCGGC	11040	
GTGTCCAA	TACATCGGTG	ACCCTTTGCA	AAACCTGGCT	AGGTGTCCAA TACATCGGTG ACCCTTTGCA AAACCTGGCT ATATTCCTGG GCCCGGTCAT	GCCCGGTCAT	11100	
AAAGTGTC	TTCATCGGCT	TCCACCCAGC	CGCCCGGATC	AGAAAGTGTC TTCATCGGCT TCCACCCAGC CGCCCGGATC CAGCATCTGT	CTGGCATAGC	11160	
CCCGTCGG	TGCCCGTCGG CCTGGTAATA CTCATCCCCT	CTCATCCCCT	ACTGCCCTCC	ACTGCCCTCC CCAAACCGCC AGATCGCCTC	AGATCGCCTC	11220	
GGATCACC	GCGGATCACC GTCCGGTTGG CCTCCGGCAT	CCTCCGGCAT	TTCACGCCGG	CTCGGCCGCT	GGATCCACCC	11280	15/6
CGCCGGTA	TTCGCAGTAA	CCCGTTGAAT	CCGCGCGCAT	CGCGCCGGTA TTCGCAGTAA CCCGTTGAAT CCGCGCGCAT GATGCACCGC TTGGGCGATC	TTGGGCGATC	11340	3
CCGGGTGG	TCACCTCGCT	TGCGCTGGCC	GCGCTGTCGC	AGCCGGGTGG TCACCTCGCT TGCGCTGGCC GCGCTGTCGC ACGGGGCGCT CGGTGGTAAC	CGGTGGTAAC	11400	
ACGTCATA	GGACGTCATA ATTAACCAGC GTAACCGAAC	GTAACCGAAC	CTAAGACCAG	CTAGCTGCGG	CAATATTGGC	11460	
CCAGGACT	GACCAGGACT ATGGCGCCCT CCGAACCCGG	CCGAACCCGG	CCGATCCATG	TCAAAACATT	GACAATGCGT	11520	
TCACGCCG	TGTCGGGCGC	GCTGAATGAC	CGCATTGCGG	ACTCACGCCG TGTCGGGCGC GCTGAATGAC CGCATTGCGG CGCTCATTCG GTGCGTAGTC	GTGCGTAGTC	11580	
TACCACCG	GCTACCACCG CAACAATGGG CTTAGGCCAT	CTTAGGCCAT	TCCTTCGTTC	ATCGCGCGG ACATGGCCGA	ACATGGCCGA	11640	
ACGCAGCG	TAACGCAGCG GTCAGCTGCT CGCCCGCCGC	ລອລລອລລລອລ	GTCGTTATAC	GTCGTTATAC GCGGACGCCG CGGCCTGCGC	CGGCCTGCGC	11700	

		FIG	FIGURE 1-16		Page 1	16 of 22
ATTGTGCAGC GCCTCGTTGA	GCCTCGTTGA	CCCGCTGAGC	CCCGCTGAGC CGCCGCCTCG GCACCCAGCT	GCACCCAGCT	TCTTCAGCAA	11760
ACCATCTTCG ATGCGCAGGC	ATGCGCAGGC	CGGTGAGCCA	CGGTGAGCCA CTGGTGCCCA TTGATCGTCA		CTTCGACGGT	11820
CTCGGCTTCG TCGGTGGCGC	TCGGTGGCGC	GGAAGGATCC	GTTGTTCATC	TGATTGAGCG	TCCCGTCTAG	11880
GGCCGACTGA AACCGCGCCG	AACCGCGCCG	CCAGCGTCAA	ອນອອອນນນອນ	ACATGCGGGT	CCAATTCGTC	11940
CATGCTCACT TCGACTCCTT	TCGACTCCTT	ACTGTCCTGG	ACTGTCCTGG CGCCGACGGT TACCAATGAC GGCCTCGGTC	TACCAATGAC	GGCCTCGGTC	12000
CATGCCCGAT CCTCGGTGTA	CCTCGGTGTA	GAGCGCCTCG	TCTTCCTGCT GAGAACCCTT GGACTTGGCG	GAGAACCCTT	GGACTTGGCG	12060
CCCCTTGTC CCTGATGCGC	CCTGATGCGC	GGCACCCATC	GGCATTCCCA	TGCCACCGCC	GCCCAGCGCG	12120
GCGCCGCCGC CGGCCCTTCC	CGGCCCTTCC	CTGGCCTAAG	CTGGCCTAAG CCGGCAATGT CACCAGCGCC AGCGGGCCGC	CACCAGCGCC	AGCGGGCCGC	12180
ACCGATTCGG CGCCCCCGAT	CGCCCCCGAT	CGCGGATCCC	CGCGGATCCC AACGGCGCCG ACGGCACCCC GCCGCCTCCA	ACGGCACCCC	GCCGCCTCCA	12240
CCGCCACCGA GCGATGCCGC	GCGATGCCGC	TTTGACCGCC	ACGTCGCCCG	ACGICGCCCG ACAGCGCIGC GGCIICCCGC	GGCTTCCCGC	12300
CCAGCCGACG TCAGCTGCGC	TCAGCTGCGC	CGCCGTGTCA	CGCCGTGTCA GCCGGGAGGC CACCACCCGG CGATCCGGTA	CACCACCCGG	CGATCCGGTA	12360
GGCGGAACCA TCGGTGCGGC	TCGGTGCGGC	TGGCATCCCG	TGGCATCCCG GTACCGGGAG TCACACCGGA GCCGTCAGAC	TCACACCGGA	GCCGTCAGAC	12420
GGCGGCATCA GGAAGCCAGG	GGAAGCCAGG	GATCAATCCC	GATCAATCCC TGCTCTTGCG GAGGCGGGC GGGTCGATCT	GAGGCGGGC	GGGTCGATCT	12480

						18/6	3						
18 of 22	13320	13380	13440	13500	13560	13620	13680	13740	13800	13860	13920	13980	14040
Page	CCACCGTCAC GGTCTGCGAC	ATCGTCGCTA AACTGAAAGG	GCGAACGGAT CGCCCTACTT	ACCCGGCGAA TCGGCAGCTG	ACCCGGTCAA TACCGGGATG	ACCTCCTGCC CGATGTGTGC	GTGACCATCG CACCTGCCGC	CCCAGGTTGA GGGCGATGTC	TCGGTTCACC GTCTCGCCGA CCAGTACCCC	AGATGCTGGC CTTGCAGCGC	GTCTTGGGCT TGTCCGCCGT	ACCCGGACTC TGGTGATGGT	GGCGCCGCGT AGGCGGCAGT
FIGURE 1-18	SCCTCCACCT CGTTGGCCCT GTTCAAAATC TCTTGCTGAT CCACCGTCAC GGTCTGCGAC	rgcgrcatat cggarcarcc rccrtagrgc ratagccarr	FTCCTGCACT AATTTGATGC CGCCCGTTCA TGCCGGCATC	EGGCAGCGCC ATCTGGTAGC GGCTTTCCTC GGGTGGGGAA ACCCGGCGAA TCGGCAGCTG	CCGATGCCGC GGGGTACCGA TCACATTGTG CCGCAGAATC ACCCGGTCAA TACCGGGATG	SGGGCCGAGA TAGGTCGTCG CATTCGGCCA CGCCACCTTT	SCCGATCAAC CGGGCAAATT CCTCGAACTG TGGCCCGACT	SGCCGCACGC ACCACGAACT GGGTGAATGT CTGAGCGTCA CCCAGGTTGA GGGCGATGTC	SACATCGTCG AAGGGCATGT AGACCGGGCA TCGGTTCACC	AGCTGACCCG ATCGGCAGCT GGCAGTGGCG GTTGGCCACC	3GGCCGCTGC CCGCCAAATA GGCGGGCGAA GCCCCTGGGT GTCTTGGGCT TGTCCGCCGT	SGTCAGCAAC ACCGTGGACT GCGGGGCCAT CCCCGGCGCG ACCCGGACTC	STGGTCCGCG CGCCCCGACC ACCATACATC CGGACCTCCG
	3CCTCCACCT	rgcgtcatat	PTCCTGCACT	JGGCAGCGCC	CGATGCCGC	regeceaaa	SCCGATCAAC	GCCGCACGC	SACATCGTCG	AGCTGACCCG	3GGCCGCTGC	3GTCAGCAAC	FIGGICCGCG

						19/63	3						
19 of 22	14100	14160	14220	14280	14340	14400	14460	14520	14580	14640	14700	14760	14820
Page	ACAAAGCCGA TGTCGGTGGC	ACACCATGGC TAGCCAGTCG	AGTCCTGCGA CGCCTTCATC	CGCAGCACGA TCCAGGTCCG	FACAGACTCA CCACGTCCGG	TGAGACGATA TCGGCCTCCA AGTCGGGACA	ATGTCGTCGG TGTGGGCTTG	AGCTCGATTA CGGCGACCAG	TTCACGGTGG CCCCGACCAC	AACCACGCGT ACACCGCCGC	SCCACGATGA CGCCCAACGA	AATGCGATGA TGCATGGCGG	GGTGCTGAAC CGCAGCCCTA AAGGATTTCT
FIGURE 1-19	GTAGGCATCG CGCCCCTTGA TCATCGACCA TTTCTCCCGC ACAAAGCCGA TGTCGGTGGC	GTGGTCGTAG TCATCGAAGC TGCGGCCACA CACCGCGTCG	ATCGGCAATG CGCGTCGCGG ACGCCACCAA ATACCGGGCC	GCGGCGCTGC GCCGATTTGC GGGTGCGTTC CGGGTCGGCG	GCGGTTCGCC GGCGCCGGGT CTGTCCCGAT CACCTGCTGA TACAGACTCA CCACGTCCGG	CGCTGCGGTA TTGCCGACGC GGTAGCCGGC TGAGACGATA 1	GTGCACCGAC AGGAGCTCCT CCACCAGTCC GGTGTCCAGC A	CCCGTCGACG ATGACCGTCG GCGTGAATGG TCGGGGAATG AGCTCGATTA CGGCGACCAG	AAACTCGCCT TGCCAGCGCA CCGCAACGTG ATCTCCTGGC TTCACGGTGG	AGGTTCTGAC GAGGAATCCG GGGGCCGTCG GCGCCGCCGC A	CACCCAGCCG GTGATCCGGC GGCCGTAGAA AGTGACCGTG GCCACGATGA CGCCCAACGA	GGCCAGCGCA ATCCCCGCCC ACCAGTAGCG CGTCTCCAAG AATGCGATGA TGCATGGCGG	GGCCAACGCG GAGGCAAGCA AGGCGTGCCC GGTGCTGAAC C
	GTAGGCAT	GTGGTCGT	ATCGGCAA	CCGCCGCT	GCGGTTCG	CGCTGCGG	GTGCACCG	CCCGTCGA	AAACTCGC	AGGTTCTG.	CACCCAGC	GGCCAGCG	GGCCAACG

						20/6	3						
20 of 22	14880	14940	15000	15060	15120	15180	15240	15300	15360	15420	15480	15540	15600
Page	CCAGGGCCAA CGTAAGGCCG	GACCCGGCTC CACCACCGGG	CGGGCGGAAT GTCCCACGTC	CGICGACCCC GCCCCCGGGG	ATCTGCGCCG GCGTCAGGTC GGGGAACCGC	TATGCCGCGG CAAACGAGGT GCCGGCGATG	CACCGGTGTC GCCGAGCGCG	CGTGCATCGA GAACGAGCTG	CTTAACACCA GCGGTGCGTA CCACGCCGGG	CGGGTGTGGA CGGGTCCGGC	TTGCCGGCCG CGACCACCAC CACCACGCCT	TTTCATCGAT CGGCCTGCTC	CGCCGAGGTT GGCGGCGTGC
FIGURE 1-20	CATCGGCGGC TCAGCGCCCG TCTAGCCAGC GCGCCCAGGC CCAGGGCCAA CGTAAGGCCG	ACGGCCACCA ACGCCACAGC CGTAATCGGG CGACGATCGG GACCCGGCTC CACCACCGGG	GGTGGAAGTC GTCTGACGTT GTATGGCGCC GAAGCAGGGC CGGGCGGAAT	AGCGCGCCA CCGCATCGAT GACGCCGGCG CCGACCAGGT	TGTCTCGCGG TGGCGGTGAT CCGGTGGATG ATCTGCGCCG	TGCCGAAGCA GGGCCGCCAG ACCCGACACA TATGCCGCGG	GGTACCGGCC CCTCCCGGCC TTGCAACGCA TTCACCGGTT (	ACGATGTTTT CTGCGGGCGC GGCCACGTCC ACCCACGGTC CGTGCATCGA GAACGAGCTG	GGCATCCCGG TCTGGCCGAT ACCGCCGACG CTTAACACCA G	GTGACAACGG TCTGCACATT GTTCCAGCCG CGTGGGTCGC (	GCCGGATTCT GTACGCAATC GCCACCGGTG TTGCCGGCCG	TTGACGTTGA CCGCATAGTC GATGGATGCA CCCAGTGAGG 1	ACCTTGTAGC AGGCGGCTTC ACTGATGTTG ATCACACCCA C

		FIC	FIGURE 1-21		Page 21 of	1 of 22
ACCACGGCGC	ACCACGCGC GGGCAAGACT	GCGGATGGAA	GCGGATGGAA CCGGCGGCCG	GGGTGGCGTT	GGGGTCATTC	15660
GGGTTGGCTT	GGGTTGGCTT GTGAGCCGAC	CGGTTCGAAG	CGGTTCGAAG GCCTCAGACG	TCTGACGTAG	CGAGAGCAGT	15720
CGAGCGTCGG	CGAGCGTCGG GCGCGACGCC	GACGAACCCG	GACGAACCCG TCGGTGGGCG CGGGCCGGCC CGCGATGATG	ວລອອລລອອອລ	CGCGATGATG	15780
GATGCTGTGA	GATGCTGTGA GAGTCCCATG	GGCATCACAG	TCAGACAGGC	CGTTACCGGC CTGGTCGACG	CTGGTCGACG	15840
AAATCGCCGC	AAATCGCCGC CAGGTTCCGC	CGGGACCCGT	GGCGAAGCGT	CGACACCGGT	GTCGATCACC	15900
GCCACCGTCA	GCCACCGICA CCCCGGCCCC	GGTCGCGAAC	TTGTGGGCAT CGGCCACGCC		CAGATACGTG	15960
TTGCTCCACG	TTGCTCCACG GCGGATCGTG	GAACCCGGAC	GAACCCGGAC CCCGGCAGCG TGGTGGGCGA CGCGCACAAA	TGGTGGGCGA	CGCGCACAAA	16020
ACGCGCTGTT	ACGCGCTGTT CGGTAGGCTG	ATCCGGGCCC	Arceggeee gecaegregg	GCGGCAACGC GCCCGGATCG	GCCCGGATCG	16080
ATCGGCGGTG	ATCGGCGGTG GCGTGATGGC	CGATGCGGGC	GACGCGGTGA	GCAACGCCAG	CGCCACCGTG	16140
ATCAGAAAGA	ATCAGAAAGA TACGGTGCAC	TCCCAGAACA	TCCCAGAACA CTCCATTCGT TGAGATTCAT TGCGATTCAT	TGAGATTCAT	TGCGATTCAT	16200
TGAGCTGCGT	TGAGCTGCGT TGCTACCTTG	GGCCACTTGA CGGACCTGTG		TGCATTTTAG ACGTAACGGC	ACGTAACGGC	16260
TGGGCAAACA	TGGGCAAACA ACGCTGTCAC	GCCTGGGCTG	GTCCGCCGCG	CCGACCAGGG	CGCGTAGGCG	16320
CTGTACCTGG	CTGTACCTGG ACCACGCCGG	GACTCAACGG	TTTTGCTACC GCACTAGCCG		ATATGCGGCT	16380

						22/6	3		
22 of 22	16440	16500	16560	16620	16680	16740	16800	16860	16885
Page	CACGCTGCGT ATCGCGGCAT	TACCGGTTGG CCCGCACGTT	TTCGATCACG TTGCGGGCGA	AGGGGTGGTG TAGTTACGGA	ATCGAGCTGG CTGGCGCGCG	CGAATAAGAC TCGAACCGCA	GGTTCACGGT GAGGAATTCG GTACCCCGGG	TTGGATGCAT AGCTTGAGTA	
FIGURE 1-22	GCTACCAAAC GATCGCGGCC ATGTCTCGGT TGTCTGAGCA CACGCTGCGT	CGATGICGGI GGCGGTGATG ATCTGCAGAT CCTGAACCGA	TTTGCGCAAC CACCCGGGTG TCCCGGAACC CTTCGGCGCG	ACCGACCGTT TTGCATAGCG TCGATACCGT GCTGCCCACT AGGGGTGGTG	TGGTGGTGAC CGCGTCGAGG AATACCTCCC GTGCGGCGTC ATCGAGCTGG CTGGCGCGCG	GTGTAGCGTA GCGGTGTCCA ATCTCGACGA TCTCCACCGG CGAATAAGAC TCGAACCGCA	GCTTTCGGTT GAACCGGCCA GCCAAACCCG GGTTCACGGT	TTCGAAATCG ATAACTTGGA TCCGGAGAGC TCCCAACGCG	TTCTATAGTG TCACCTAAAT ACTTG

		FIC	FIGURE 2-1		Page 1	of 20	
GGATCCTCGG ACTGGCCGCG GTCGTGCTTG	TGGCCGCG	GTCGTGCTTG	TGCACGAGTT	TGCACGAGTT CACCGAGGTC ATCGTCATCG	ATCGTCATCG	09	
CCAACGGCGT GCGGGCCGGA CGCATCAAAC	GGGCCGGA	CGCATCAAAC	CACTTGCCGG	CACTTGCCGG GCCACCCAAG ACACCTGATC	ACACCTGATC	120	
GGACTATCCC GGGGTAGCGA CGCGCGGAAT CGTGGAGTGT GTTTGGACCA GCAATAGCGT	GGTAGCGA	CGCGCGGAAT	CGTGGAGTGT	GTTTGGACCA	GCAATAGCGT	180	
CACTGTGACG AAACAGCCGC CGTCTTCTGG AAGTTATACC CGGTTATACT ATCTGTATGA	ACAGCCGC	CGTCTTCTGG	AAGTTATACC	CGGTTATACT	ATCTGTATGA	240	
AGACAGCTAT TTCTCTGCCG GATGAGACGT	CTCTGCCG	GATGAGACGT	TCGATCGGGT	TCGATCGGGT ATCGCGGCGT GCGAGTGAGC	GCGAGTGAGC	300	
TCGGCATGAG TCGGTCCGAG TTCTTCACGA	GGTCCGAG	TTCTTCACGA	AGGCTGCGCA	AGGCTGCGCA GCGCTACCTG	CACGAGCTGG	360	23/
ACGCCCAATT GCTCACGGGC CAGATCGACA GGGCTCTAGA GAGCATCCAT GGCACCGACG	TCACGGGC	CAGATCGACA	GGGCTCTAGA	GAGCATCCAT	GGCACCGACG	420	63
AAGCGGAGGC CCTCGCCGTG GCCAACGCAT ACCGCGTGCT AGAAACCATG GACGATGAGT	TCGCCGTG	GCCAACGCAT	ACCGCGTGCT	AGAAACCATG	GACGATGAGT	480	
GGTGATTAGT CGTGCCGAGA TCTACTGGGC TGACCTCGGG CCGCCATCAG GCAGTCAGCC	TGCCGAGA	TCTACTGGGC	TGACCTCGGG	CCGCCATCAG	GCAGTCAGCC	540	
GGCGAAGCGC CGCCCGGTGC TCGTAATCCA	CCCGGTGC	TCGTAATCCA	GTCAGATCCG	TACAACGCAA GTCGCCTTGC	GTCGCCTTGC	009	
CACTGTGATC GCAGCGGTGA TCACGTCCAA TACGGCGCTG GCGGCAATGC CCGGCAACGT	AGCGGTGA	TCACGTCCAA	TACGGCGCTG	GCGGCAATGC	CCGGCAACGT	099	
GITCITGCCC GCGACCACAA CGCGACTGCC ACGTGACTCG GTCGTCAACG TCACGGCGAT	GACCACAA	CGCGACTGCC	ACGTGACTCG	GTCGTCAACG	TCACGGCGAT	720	
TGTCACGCTC AACAAGACTG ACCTCACCGA CCGAGTTGGG GAGGTGCCAG CGAGCTTGAT	CAAGACTG	ACCTCACCGA	CCGAGTTGGG	GAGGTGCCAG	CGAGCTTGAT	780	

24/63
-------

	FIGURE 2-2	Page	2 of 20
GCACGAGGTT GACCGAGGAC TTCGTCC	SCGT ACTGGACCTT	TTCGTCGCGT ACTGGACCTT TGACACTGCG CCACGCGACA	840
ATTCGTCACG GTGACGTTCC TGCTTGC	STGT AAGCCCCCCC	TGCTTGGTGT AAGCCCCCCC GCCGGGGGAA CTACTCGCCG	006
GAGGTGTT TGTGGGCAGG CTTGAGGGCA	GCA AGGTTGCATT	CATTACGGGC GTGGCTCGGG	096
GTCAAGGCCG TTCGCATGCG GTCCGCCTAG		CCGACGGCCA AGCGCGTGCG CTCGGCAAGG	1020
TCGATGTTGA GGCGTGCGGT GCGCTCGTTG	STTG GTGAGGTAGA	GTGAGGTAGA AGTGTGGGGC CGTGACGTGC	1080
GTGACGATCG ACGGGTGTTT GTCGAGAGTC		CTGCCGACGA GTTCGGCGCG TGCCGCCGCG	1140
TCGCGCGTCA GGGCATCCGT GTCGTAGGGC	GGC TGCCCGTTTC	ACAGAGGAA CTTGTCGAGC	1200
CCGAAGCCGG GTGCGCGGCG AGGCGCT	CGG CTGCTGGCTC	AGGCGCTCGG CTGCTGGCTC CCAGTAGACA TCTAGGCCTG	1260
CGTCGACTGC GGCTGCGGCA GCGTCGTGCT	GCT GGTGACGAGT	GGCGTTGGTG TCCAGCGTGA	1320
TCGCAGTGGT GCCGGCGTGG TCGCGGGACA	ACA GGAAGTCCTC GACCGGTTTG	GACCGGTTTG TGATCACCCG	1380
GCCCGAGCCG AAACTGAATG CCCATCGTCG	TCG TGAAGTTCCT	TGAAGTTCCT CTCGCATCGA CGCCTCGGTT	1440
CGTGTCATAA TACATGACAA ATCAATAGAC		AAAAGGAAGA CAGGCTGCCC ATGGGAGTAA	1500
ATGTGCTCGC CTCGACCGTG TCGGGTGCGA	CGA TCGAGCGCTT	GGGATTGACC TACGAGGAAG	1560

<u> </u>
age
Ба
щ
i
•
(1
$\Xi$
FIGURE ?
$\succeq$
豆
<b>—</b>

		FIC	FIGURE 2-3		Page 3	of 20	
TCGGTGACAT	TCGGTGACAT CGTCGATGCC	TCGCCGCGTT		CCGTGGCGCG ATGGACCGCA GGTCAGGTGG	GGTCAGGTGG	1620	
TTCCCCAACG	TTCCCCAACG CCTCAACAAG	CAACGACTTA	TCGAGCTGGC	CTATGTCGCC	GACGCCCTCG	1680	
CGGAAGTGCT	CGGAAGTGCT GCCGCGTGAC	CAGGCGAACG	TGTGGATGTT	CAGGCGAACG TGTGGATGTT TTCGCCGAAT CGGTTACTGG	CGGTTACTGG	1740	
AACACCGCAA	AACACCGCAA GCCTGCCGAC	CTCGTGCGAG	ACGGCGAGTA	CTCGTGCGAG ACGGCGAGTA CCAACGCGTG	TTGGCGCTCA	1800	
TCGACGCGAT	TCGACGCGAT GGCGGAGGGA	GTGTTCGTGT	GAGCGATGCC	CTCGATGAAG	GGCTCGTCCA	1860	
GCGTATCGAC GCACGCGGAA	GCACGCGGAA	CAATTGAGTG	GTCGGAAACG	TGCTACCGGT	ATACCGGCGC	1920	
GCACCGTGAC GCCTTGTCCG	GCCTTGTCCG	GTGAGGGCGC	GCGCAGATTC	GTGAGGGCGC GCGCAGATTC GGAGGCAGGT GGAATCCGCC	GGAATCCGCC	1980	
GCTGCTCTTT CCGGCGATCT	CCGGCGATCT	ATCTTGCTGA	TTCCGCCCAA	ATCTTGCTGA TTCCGCCCAA GCCTGCATGG TTGAGGTGGA	TTGAGGTGGA	2040	
ACGGCGCC CAAGCGGCTT	CAAGCGGCTT	CAACGACCGC	AGAGAAGATG	CTCGAGGCGG	CCTACCGACT	2100	
ACACACGATC GACGTCACGG	GACGTCACGG	ACCTGGCCGT	CCTCGATCTG	ACAACCCCGC	AAGCTCGGGA	2160	
AGCCGTGGGG CTCGAGAACG	CTCGAGAACG	ACGACATCTA	TGGCGACGAC	ACGACATCTA TGGCGACGAC TGGTCAGGGT GCCAGGCGGT	GCCAGGCGGT	2220	
CGGACATGCG GCCTGGTTCT	GCCTGGTTCT	TGCACATGCA	AGGTGTCCTC	TGCACATGCA AGGTGTCCTC GTGCCGGCGG CGGGCGGTGT	CGGCGGTGT	2280	
CGGCCTCGTT GTCACCGCGT	GTCACCGCGT	ATGAACAGCG	AACTCGGCCG	GGCCAACTAC AACTGCGACA	AACTGCGACA	2340	

						26/6	33						
of 20	2400	2460	2520	2580	2640	2700	2760	2820	2880	2940	3000	3060	3120
Page 4	TAGC TGGCCAGCTT	CTCG CATCACTTAT	AAGCCGATAA GCGACATTAT GTCAAGTGAA GCTGGTCGTA TTGGGTTAGC TGCGCCGTTT	GTGCTAGCGG GCACGCTCCT TGTGCGTGCT GCGGCAGCGA GCGTGTCGTC AAAGGTTGCG	GTAG TTTTAGCCCG	ACGG CGCAACGACG	CTACCCGCGG AACGCGGATC GTCGAGCATC GATGATCAGG CGCCGCAATA ATTTGGAACG	ACCT CCGGGGTTTC	CCAC TTCTAGCAAA	GGTGGTACAC CTACTGGCGG CCGCGGGTTT ACCGCCCCTG CCAGTCACCG CACTTCCGGC	CGCG CGCCGTGGCA	SAAT TTACATACGA	GGGCTCGCTG CAGCACGAGA TCGGCCGCGA CAAGCCCGCT GTTGGGTGCC AGCAGCGTCC
	GCCACG	GAGCAACTCG	TTGGGT	GCGTGT	AATCAGGTAG	CGTCGTACGG	ລອລລອລ	TGAGCT	AGTGTGCCAC	CCAGTC	CAGGAA	AGAAGCGAAT	GTTGGG
FIGURE 2-4	AGAACTTCGA	TCATCGTAAG	GCTGGTCGTA	GCGGCAGCGA	GGCACGCAGC	AGCTCGCCAT	GATGATCAGG	GTCCAATGCC	CATAGAACGA	ACCGCCCCTG	TCCACAT GCGGCGGTGG CAGGAACGCG	GCGCGTTGCG	CAAGCCCGCT
FIC	CTCTTTACCA	TGTGCCATGG	GTCAAGTGAA	TGTGCGTGCT	TGCCACATCC	GAGACGCAGC	GTCGAGCATC	eeeceeccce	GAGTGGTCGA	CCGCGGGTTT		GTGAACCTTC	TCGGCCGCGA
	AAGCGTCGAT CTGACGCCTG CTCTTTACCA AGAACTTCGA GCCACGTAGC	GGCGCAGAGA AGGATGCCGC TGTGCCATGG	GCGACATTAT	GCACGCTCCT	AGGCTTGCCT GGTGATGAAT TGCCACATCC	CTGCTAGCGC GTAGTTCGGC GAGACGCAGC AGCTCGCCAT	AACGCGGATC	GGGCTCGCCA GGCCATCGCT GGGCGGCCCG GTCCAATGCC TGAGCTACCT	GGTTATTTGG TAGCGCGGAC GAG	CTACTGGCGG	GGCGGACGAC GAGCACCGCG GAA	GCATGAGGCC GACGGTCAGG GTG	CAGCACGAGA
	AAGCGTCGAT	GGCGCAGAGA	AAGCCGATAA	GTGCTAGCGG	AGGCTTGCCT	CTGCTAGCGC	CTACCCGCGG	GGGCTCGCCA	GGTTATTTGG	GGTGGTACAC	GGCGGACGAC	GCATGAGGCC	GGGCTCGCTG

		27/63

FIGURE 2-5		Page 5 c	of 20
GCAGCAGGCG GGACGAAATC CCGTACGGCG GGTTCGCCAC AACCCGGAAC GGCCGGCCGG	ACCCGGAAC G	ອອລລອອລລອຄ	3180
GCAACCGGAT CGAGGCGGCG TCCGCGTGCA CCACGGTAAT G	GCCAGGGAAT C	CGCTCGCGGA	3240
GGACACCGAC TCGTCGCGGG TGCAACTCCA CGGCGACCAC CCGCGCCCCC GCTCGCACTA	ם בפכפכככככ	CTCGCACTA	3300
GATGCGCCGT CAGTGCCCCT TCGCCGGCGC CGATGTCAAA CACGAGCTCA CCGGACCGCA	CACGAGCTCA	CGGACCGCA	3360
CTGCGGCCGC GCTGACTACC CGCGCTGCCC ATTCGTCATG G	GAGCCGGTGC CAGCCCCATG	AGCCCCATG	3420
CCCGTCGCGA CCGTCCGAGG GCGGACACGA CGTACCGTCA C	CTGCGTAGAT G	GCCCACGCGC	3480
CCGACCGTAG CCCGCCACCG GCACTGCGAT CAATCCAATT TCTCGGTTCA GGCAACCTTC	CTCGGTTCA G	GCAACCTTC	3540
TGGTCATCAC CAGCCCCAGG GCTCTGGCGC CGTCCGCATC AACTCCGAGA TGACGTTGGC	ACTCCGAGA T	GACGTTGGC	3600
CGTGACGACC CACTAGACCC ACCTGGCAGT AGCCGCATTG T	TCGCAGTCGG C	CGAGCCTCAG	3660
TGCGCAGTCG CGTCTAGGTG CAAGGATATT GCCCGTTGAG C	CAGACAACTC GACGGCGGCG	ACGGCGGCG	3720
AGTAAGAACC GGTCAGCCCG CCTCTTAGGC CGCCCGTGGC TGAACCACCG GGGGCAATGA	GAACCACCG G	GGGCAATGA	3780
TGCGATTCCA ATTCGCTGGG CTGAGAACGT AGTGCGTGCC AGATCGTGCA ACGGTGCTAT	GATCGTGCA A		3840
TCCATGTGTG CAAGACGGAT TCTCCTGCCG GCAAGTCGAA T	TTCAAGCTTC C	CAATCGGTTA	3900

v
Page
g d
(A
压 2-6
闰
5
FIGURE
E

		FIGURE 2-6		Page 6	6 of 20
GCGCCCCT GCTCGAGTTT G	TGATGGTGA	AGCGGGCGAT	GTGATGGTGA AGCGGGCGAT GAAACCGGTC TGCCACGTCG	TGCCACGTCG	3960
ATGTCACCGA CAACGTCGCC C	TGGCCGTCG	CCGCACTAGC	CTGGCCGTCG CCGCACTAGC GACCGGGGTG ATGGCGAGTC	ATGGCGAGTC	4020
CGAGGATGGC AACTATCAAT GO	CCGACACGG	TTGCGTGAAG	GCCGACACGG TTGCGTGAAG CGCTGTCCGC CAGCGCCTCA	CAGCGCCTCA	4080
CGTAAATGTT CAGTCCGGCC A	TGACAGCCA	ACACTAATGC	ATGACAGCCA ACACTAATGC CAATGAGGCG ATATCGGCCG	ATATCGGCCG	4140
TCTCCTCGCG AGCAAGCTAC AG	AGCAACTTTG	CTCAACCGCA	CTCAACCGCA ACCGTGATGA AATTTGGCCT	AATTTGGCCT	4200
CGACCCACCC TGAACCAGAT AT	TCGGCCCGG	CCGAACGCGA	ATCGGCCCGG CCGAACGCGA ACTTGCGGAC GGGGAAGGCC	GGGGAAGGCC	4260
AGACAGCCTC GACCCCACTC CO	CCCGATTAG	CGCCGTTCAC	CCCCGATTAG CGCCGTTCAC CGTTCGCGAC CGGTATCAAC	CGGTATCAAC	4320
GGGCTACAGC TCCAACACGA TO	TCCGTAGGGC	CGCGTCACGC	CGCGTCACGC CGAATGTGCA CTGGTGGCGC	CTGGTGGCGC	4380
CGACACGCCC GGGCGAGGCC GG	GCCGTCGGCG	TGTCAGCTGG	TGACTGAGTT GTGCAGACTG	GTGCAGACTG	4440
ACCGCGCGC CTCCTGCCGA AC	CGGTATGTG	CCCATCGACG	ACGGTATGTG CCCATCGACG ATCACGTGGT CCAACCCGCG	CCAACCCGCG	4500
TGTGCACACG TGCTGTACTA GC	GTCACGGTC	AGCGAGATTC	GGTCACGGTC AGCGAGATTC CCAGCGCAAC CATCATGACC	CATCATGACC	4560
GCGATCAGGC CGTCGAGGAT TO	TCTCCACGAG	CCGGGGTTGG	TGAACAGCCC	GCGCAACCGG	4620
CCGGCTCCGA ACCCGAGGGT GC	GCGAACCAT	ACCGCACTGG	GGCGAACCAT ACCGCACTGG CTGTGACCGC GCCGAGGCCG	GCCGAGGCCG	4680

29	6:

FIGURE 2-7	Page 7 of 20
AACAGCCAGC GCTGGTCGCT GTGCTCGTTG GCCAGCGCGC CTAGCAACAC GACGGTGTCG	GGTGTCG 4740
AGGTAGACGT GTGGGTTGAG GAACGTGAAT GCCGCACAGG TCACCAGGAC CTC	CTCGGCTAAG 4800
CGAACCGGCG TGGCGCCAGA TGGGATCAGC GCAACAGGTC GCCACGCCCG CCG	CCGGGCCGCA 4860
AGTAGCCCGT AGCCGATTAG GAAGGCGGCG CCGCCAAACT TGACGACATT GAG	GAGCGCACGC 4920
GGATGTGCGC CGATCAATGC GCCGAACCCC GCGATACCGG CGGCGATCAG CACGATGTCG	GATGTCG 4980
GACACCGTGC ACAGCGCCAC CACCGGCAGC ACGTGCTCAC GCTGGATTCC CTG	CTGCCGCAGC 5040
ACGAATGCGT TCTGCGCGCC AATCGCGGCG ATCAGCGTGA AGCAGGCCAG GAA	GAAGCCGACG 5100
ACCAGTGGTG AGTTCACGCA ATCGACACTA GGCAGTTTGT ATGGGTCAGT ATAGCTAATA	GCTAATA 5160
ATTCTTCATT TACATTAGCA TTATTAATGT GCAGTGCGAC GCTCCGCAGA TGG	TGGTCTACAC 5220
CTGAGATGGT GGATCCGCAG CTTGACGGTC CACAGCTGGC CGCATTGGCT GCC	GCCGTGGTCG 5280
AACTGGGCAG CTTCGATGCG GCCGCGGAGC GCCTACATGT CACCCCCTCG GCTGTCAGTC	GTCAGTC 5340
AGCGCATCAA GTCGTTGGAG CAGCAGGTCG GCCAGGTGCT GGTGGTCAGG GAAAAGCCAT	AAGCCAT 5400
GTCGGGCGAC GACCGCAGGT ATCCCGCTGT TGCGGTTGGC CGCGCAAACA GCGTTGCTCG	Trgcrcg 5460

FIGURE 2-8	Page 8 of 20
AGTCCGAGGC GCTCGCTGAA ATGGGTGGCA ACGCGTCGCT GAAACGCACG	ACGCACG CGGATCACCA 5520
TTGCGGTAAA CGCCGATTCC ATGGCGACAT GGTTTTCGGC CGTG	CGTGTTCGAC GGTCTCGGCG 5580
ACGICCIGCI CGACGIICGG AICGAGGACC AGGACCAIIC CGCGCGGCIG CIACGGGAGG	GGCTG CTACGGGAGG 5640
GTGTGGCGAT GGGCGCGGTG ACCACCGAGC GGAACCCGGGT GCCGGGCTGC CGGGTGCACC	GGCTGC CGGGTGCACC 5700
CGCTGGGTGA AATGCGCTAC CTACCAGTGG CCAGCAGGCC ATTC	ATTCGTCCAG CGCCATCTAT 5760
CCGACGGGTT CACTGCCGCC GCGGCGGCTA AAGCTCCGTC ACTG	ACTGGCGTGG AATCGTGACG 5820
ATGGGCTGCA GGACATGTTG GTGCGTAAGG CCTTTCGTCG CGCCATCACC AGACCGACGC	ATCACC AGACCGACGC 5880
ACTITIGICCC GACCACAGAG GGCTICACCG CCGCAGCGCG CGCC	CGCCGGGCTG GGATGGGGCA 5940
TGTTCCCCGA GAAGCTGGCA GCATCTCCGC TTGCCGATGG ATCG	ATCGTTCGTA CGGGTCTGCG 6000
ACATACACCT CGACGTCCCT CTCTATTGGC AATGCTGGAA ACTGGACAGT	GACAGT CCGATCATCG 6060
CGCGAATTAC CGACACGGTG AGGGCGGCGG CAAGCGGTCT GTACCGGGGC CAGCAACGCC	CGGGGC CAGCAACGCC 6120
GCCGCCGACC GGGTTGACCG ACGCCAGCAT GTTGTTGTGT CAGCGCGCT	GCGGCT TGGTCTCGAT 6180
GICCCGGCCT IGCIGGACCC GCTICCTCAA ACAGGTIGAA CITA	CTTAACGACT CAGACGGAAA 6240

WO 96/25519 PCT/US96/01938

FIGURE 2-9	age 9	Page 9 of 20
CGCTTGAACC GCGACGTCGC TCCGGACACC AATTTGACTC GGCTCTTTGG CAATTGAAGG	<b>A</b> GG	6300
TGAGCTGCGA GCAGCCGGGT GACCGCATCG TTGGCCTTGC CATCAATCGC CGGCTCGCGG	566	6360
ACGTAGATAA TCAGCTCACC GTTGGGACCG ACCTCGACCA GGGGTCCTTT GTGACTGCCG	SCG	6420
GGCTTGACGC GGACGACCAC AGAGTCGGTC ATCGCCTAAG GCTACCGTTC TGACCTGGGG	366	6480
CTGCGTGGGC GCCGACGACG TGAGGCACGT CATGTCTCAG CGGCCCACCG CCACCTCGGT	3GT	6540
CGCCGGCAGT ATGTCAGCAT GTGCAGATGA CTCCACGCAG CCTTGTTCGC ATCGTTGGTG	3TG	0099
TCGTGGTTGC GACGACCTTG GCGCTGGTGA GCGCACCCGC CGGCGGTCGT GCCGCGCATG	\TG	0999
CGGATCCGTG TTCGGACATC GCGGTCGTTT TCGCTCGCGG CACGCATCAG GCTTCTGGTC	3TC	6720
TTGGCGACGT CGGTGAGGCG TTCGTCGACT CGCTTACCTC GCAAGTTGGC GGGCGGTCGA	GA	6.780
TTGGGGTCTA CGCGGTGAAC TACCCAGCAA GCGACGACTA CCGCGCGAGC GCGTCAAACG	₹CG	6840
GTTCCGATGA TGCGAGCGCC CACATCCAGC GCACCGTCGC CAGCTGCCCG AACACCAGGA	3GA	0069
TIGIGCTIGG IGGCTATICG CAGGGIGCGA CGGICAICGA TITGICCACC ICGGCGAIGC	၁၅:	0969
CGCCCGCGGT GGCAGATCAT GTCGCCGCTG TCGCCCTTTT CGGCGAGCCA TCCAGTGGTT	TT	7020

32	/	6	2
VL	•	v	٠

					JL/ U	•						
7080	7140	7200	7260	7320	7380	7440	7500	7560	7620	7680	7740	7800
TATAGCTCTA	AATATTATGG	GCGGCGAACA	CTGTAGTCGA	GATGAAATGA	GGGCTGGTGT	GACGGGATGG	CCACCATCGA	GCAGGTAGCG	TGGTTGAGGT	GACCCAACTC	GTTGTCGGAT	ATGGACAAGA
CGGTCCGCTG		GGCGACATTC	ACCGCTGGGG			GATGACATAT	TATGGGAGTC	GACGTTGGGC	GAGCACTCCC	CACCGCATTG	GAACGGGGTT	GGACGTTGGG
	CAATATGCAC	CAAGCCAGGC		ACCCGGTATT	CGTAGAGCCG	TTCCGGATCC	AGGGGACGGG	CGACGACCTC	GTGCGTCGAT	GGCCTGGTCC	AATAGGCTTC	
	CCCGACGATC	TCGGGGATGA	TCAAAGACTG	AAGGGCAAGA	GTGTTGAACG		TCGATGATCG	ATCGGCGATC	TCACCTGCGG	GGAATGCTGG	ACGGCACCGA	TAGGTCGAAA TGGCGAAGGG
GTTGTGGGGC	CTTGTGTGCT	GTATGTTCAG	CGCCGGATGA	CTGGAATCTG	GTAATCGTTT	GTTTGTGTTC	CCCACCGGAA	CAGGGTGGTG	GAACACGAGC	CGCCATCGTC	GTCCACGCCG	CGGCCAGCAA GGCGCTGAAG
TCTCCAGCAT	AGACCATAAA	CGCATGTTTC	GGCTCGATCA	TGTACACCGG	CGGTCGGGCG	AGACCTCAAT	TTCCCGTTAC	TCTTTACGTA	GGTTGGGACC	CACCCGGTAA	CCAGAACGCC	CGGCCAGCAA
		GGCGGGTCGT TGCCGACAAT CGGTCCGCTG TATAGCTCTA	GGCGGGTCGT TGCCGACAAT CGGTCCGCTG TATAGCTCTA CCCGACGATC CAATATGCAC CGGAGGCGGC AATATTATGG TCGGGGATGA CAAGCCAGGC GGCGACATTC GCGGCGAACA	GGCGGGTCGT TGCCGACAAT CGGTCCGCTG TATAGCTCTA CCCGACGATC CAATATATGCAC CGGAGGCGGC AATATTATGG TCGGGGATGA CAAGCCAGGC GGCGACATTC GCGGCGAACA TCGAAAGACTG TTGTCCCTAT ACCGCTGGGG CTGTAGTCGA	TGCCGACAAT CGGTCCGCTG TATAGCTCTA CAATATATGCAC CGGAGGCGGC AATATTATGG CAAGCCAGGC GGCGACATTC GCGGCGAACA TTGTCCCTAT ACCGCTGGGG CTGTAGTCGA ACCCGGTATT CATCAGGCCG GATGAAATGA	TGCCGACAAT CGGTCCGCTG TATAGCTCTA 7080 CAATATGCAC CGGAGGCGGC AATATTATGG 7140 CAAGCCAGGC GGCGACATTC GCGGCGAACA 7200 TTGTCCCTAT ACCGCTGGGG CTGTAGTCGA 7320 ACCCGGTATT CATCAGGCCG GATGAAATGA 7320 CGTAGAGCCG ATCACCGCCG GGGCTGGTG 7380	TGCCGACAAT CGGTCCGCTG TATAGCTCTA 7080 CAATATGCAC CGGAGGCGGC AATATTATGG 7140 CAAGCCAGGC GGCGACATTC GCGGCGAACA 7200 TTGTCCCTAT ACCGCTGGGG CTGTAGTCGA 7260 ACCCGGTATT CATCAGGCCG GATGAAATGA 7320 CGTAGAGCCG ATCACCGCCG GGGCTGGTGT 7380 TTCCGGATCC GATGACATAT GACGGGATGG 7440	TGCCGACAAT CGGTCCGCTG TATAGCTCTA 7080 CAATATGCAC CGGAGGCGGC AATATTATGG 7140 CAAGCCAGGC GGCGCAACA 7200 TTGTCCCTAT ACCGCTGGGG CTGTAGTCGA 7260 ACCCGGTATT CATCAGGCCG GATGAAATGA 7320 CGTAGAGCCG ATCACCGCCG GGCCTGGTGT 7380 TTCCGGATCC GATGACATAT GACGGGATGG 7440 AGGGGACGGG TATGGGAGTC CCACCATCGA 7500	TGCCGACAAT CGGTCCGCTG TATAGCTCTA 7080 CAATATGCAC CGGAGGCGC AATATTATGG 7140 CAAGCCAGGC GGCGACATTC GCGGCGAACA 7260 TTGTCCCTAT ACCGCTGGGG CTGTAGTCGA 7320 ACCCGGTATT CATCAGGCCG GATGAAATGA 7380 TTCCGGATCC GATGACATAT GACGGGATGG 7440 TTCCGGATCC GATGACATAT GACGGGATGG 7500 AGGGGACGGG TATGGGAGTC CCACCATCGA 7500 CGACGACCTC GACGTTGGC GCAGGTAGCG 7560	TGCCGACAAT CGGTCCGCTG TATAGCTCTA 7080  CAATATGCAC CGGAGGCGGC AATATTATGG 7140  CAAGCCAGGC GGCGACATTC GCGGCGAACA 7200  TTGTCCCTAT ACCGCTGGGG CTGTAGTCGA 7320  ACCCGGTATT CATCAGGCCG GATGAAATGA 7320  CGTAGAGCCG ATCACCGCCG GGGCTGGTGT 7380  TTCCGGATCC GATGACATAT GACGGGATGG 7440  AGGGGACGG TATGGGAGTC CCACCATCGA 7500  CGACGACCTC GACGTTGGC GCAGGTAGCG 7560  GTGCGTCGAT GAGCACTCC TGGTTGAGGT 7620	GCGGGTCGT TGCCGACAAT CGGTCCGCTG TATAGCTCTA 7080 CCGACGATC CAATATGCAC CGGAGGCGGC AATATTATGG 7140 CGGGGATGA CAAGCCAGGC GGCGAACAA 7200 CAAAGACTG TTGTCCCTAT ACCGCTGGGG CTGTAGTCGA 7260 AGGGCAAGA ACCCGGTATT CATCAGGCCG GATGAAATGA 7320 TGTTGAACG CGTAGAGCCG ATCACCGCCG GGGCTGGTGT 7380 CCGGCAGGG TTCCGGATCC GATGACATAT GACGGGATGG 7440 CCGGCAGGG TTCCGGATCC GATGACATAT GACGGGATGG 7500 TCGGCGATC CGACGACCTC GACGTTGGGC GCAGGTAGCG 7560 CCGCTGCGG TGCGTCGAT GACGTTGGGC 7620 GAATGCTGG GTGCTCGAT GACCTCCC TGGTTGAGGT 7620	GCGGGTCGT TGCCGACAAT CGGTCCGCTG TATAGCTCTA 7080 CCGACGATG CAATATGCAC CGGAGGCGGC AATATTATGG 7140 CGGGGATGA CAAGCCAGGC GGCGAACA 7200 CCAAAGACTG TTGTCCCTAT ACCGCTGGGG CTGTAGTCGA 7260 AGGGCAAGA ACCCGGTATT CATCAGGCCG GATGAAATGA 7320 TGTTGAACG CGTAGAGCCG ATCACCGCCG GGGCTGGTGT 7380 CCGGCAGGG TTCCGGATCC GATGACATAT GACGGGATGG 7440 CGATGATCG AGGGGACGG TATGGGAGTC CCACCATCGA 7500 TCGGCCATC GACGTTGGGC GCAGGTAGCG 7560 CACCTGCGG GTGCCTCC GACGCTCCC TGGTTGAGGT 7620 GAATGCTGG GTGCGTCGAT GACCCTCCC 7661 GAATGCTGG GGCCTGGTCC CACCGCATTG GACCCAACTC 7680 CGGCACCGA AATAGGCTTC GAACGGGTT GTTGTCGGAT 7740

WO 96/25519 PCT/US96/01938

					;	33/63							
11 of 20	7860	7920	7980	8040	8100	8160	8220	8280	8340	8400	8460	8520	8580
Page 1	CCCGAAGTCC ACCGTCGTGG	SCTGATGCTC AATCCGGTTG	CTCAGGTGAG ACAACAAGAT	GCCGTTGACG TTGGCATGTA	TCTGCCGTCG CCGCCTGGCC	CAACAGGC CGCCGCGCC TCCGGCGCCG CCCGTGCCAC	GAGGCCGC CGGCTCCGCC GTTCGCTCCC CAGCCCAAAA	SCCGGGGATG CTCGACGCGC	ACCGTTTGCC GCCGGCTCCC	CTCCCCCG GCGCCGT TGCCGATCAA CGCCGCGGAA	GTCGCGGAAT AACCGTTGCC GCCGTTGCCG	TCCCGGGTGC TCCGTCGCCG	CGCGCCCAG CAGATCGTGC
FIGURE 2-11	GGACGACATT AACGGTGGTC GGCGCGGTGA CGATGCCATT CCCGAAGTCC ACCGTCGTGG	TATACGTGGC GAAGATGTAG TACAGCCCCC CGCTGTAACC GCTGATGCTC AATCCGGTTG	CCCCGACATC	AATGGTTG ACGTCGGTCC	TATCTCCA GCGGGACGGT	TACAACAGGC CGCCGCGGCC	CCGAGGCCGC CGGCTCCGCC	CCGCCGGCTC CACCACTCAT GCCGGGGATG	GTTGCCGA GCAGCCAGCC	GCTCCCCCG GCGCCGCCAT 1	CCGACGCC	CATTCGGA CTCGTTGCCG	ATTAGCGGAC GCCCCAACAA CGTTTCGGTT GGTGCATTGA CCGCGCCCAG CAGATCGTGC
	AACGGTGGTC	GAAGATGTAG	GGAGGCCCAT GTGAAGCACT CCCAGGATTC	CAGCGGATCC GGTGTCGACC AG	CCGTCGGCTC TGTGACATGA AT	CACTGACGCC GTATCCGCCA TA	GATCGACACC GACTCCATCA CC	ACCGCCGGTT	CCCCGGGCCC GCCGATCCCG CC	CCGGGGCGT TGGGGCCGCC GG	CCGCCGGCAC CGCCGCCGAC CC	TACAGCAGCC CACCCGCCCC GC	GCCCCAACAA
	GGACGACATT	TATACGTGGC	GGAGGCCCAT	CAGCGGATCC	CCGTCGGCTC	CACTGACGCC	GATCGACACC	GTCCTGTGGC ACCGCCGGTT	ລລລອອອລລລລ	CCGGGGGGCGT	CCGCCGGCAC	TACAGCAGCC	ATTAGCGGAC

						34/6	3						
of 20	8640	8700	8760	8820	8880	8940	0006	0906	9120	9180	9240	9300	9360
e 12	บ	Æ	Æ	<b>⊲</b> :	E-i	L	4	<i>T</i> )	٠.	<u> </u>	zh		
Page	TCCTGACGC	GACCGCTTGA	ATCAGCGCC	CAGTGTGGA	GACCAGAAA	TCCAGACGCT	TAGGTAGATI	TGAGGTAGA	CATCCGGCAT	GGCACAACTT	CATCACCAAG	ACCGCAATGA	GCCACCCTGG
	CCGCATACGA TCCTGACGCC	CCTGCGCGCT	CCGATACCTC	TCGCCGCGCG	CGGGAACCAC GACCAGAAAT	ATACAGCGTA	CAGCACGCGT	TCAAAGACCC	AGACTGCCGC	GCGACCGTGA	ACCGCTCCCC	GAGTCTGAGC	GCGGCCGCAC GCCACCCTGG
FIGURE 2-12	GCCTCCGCGG	AAAGCGCTCG	ACAGCGCT GCCACCGCCG CCGATACCTC ATCAGCGCCA	AGGCGGCA GCCGCGCAT TCGCCGCGCG CAGTGTGGAA	CCGCGGAC GTCAAGAACT	CCCTGCGCGC	CCGCACGGCA	GCGCTGAATA	CACACCAGCA	TCGCGCCCG GCGACCGTGA	CTGCTTGCCG	CCCGCAACCG	TTCAAGGCGA
FI	CGCGATGGTG	CTGTTGATGG	AAACAGCGCT	CGAGGCGGCA	TGCCGCGGAC	CGATCACCAT	TCAGTAGCTA	TCTACCTGTG	TGACCGGCTG	GTTTTTGGAG	AATCTGGCCC	CGGCTACGAC	
	TGCGCGGTCT GCAACTGTGA CGCGATGGTG GCCTCCGCGG	GAGTTCAGCG TCTGCACGAA CT	TATTCCTGAC CGAACCTGGC AA	GCGGCCGCAA GCGCGGTGGT CG	CCTATGTTCT CCACATCCGC	GACACGCCGC CCCTCCGCCT CGATCACCAT	GCCTTTGACA TCTCGGATTT TCAGTAGCTA CCGCACGGCA CAGCACGCGT TAGGTAGATA	GTGGCTATTT GCTGGTACCA TCTACCTGTG GCGCTGAATA TCAAAGACCC TGAGGTAGAC	CGACTAGCCG CCGAACTCGC TGACCGGCTG	GCCCTGTCTG CCCAGCTGGC GTTTTTGGAG	CTCGACATCT TGCGTACCGA AATCTGGCCC CTGCTTGCCG ACCGCTCCCC CATCACCAAG	CTCGAGCGCG AACAAATCCT CGGCTACGAC CCCGCAACCG GAGTCTGAGC ACCGCAATGA	TCGTGGACAC AAGCGCCGTG GTGGCCCTGG
	TGCGCGGTCT	GAGTTCAGCG	TATTCCTGAC	GCGGCCGCAA	CCTATGTTCT	GACACGCCGC	GCCTTTGACA	GTGGCTATTT	CGACTAGCCG	GCCCTGTCTG	CTCGACATCT	CTCGAGCGCG	TCGTGGACAC

മ
Page
Б
•
3
2-1
d
5-3
$\Xi$
5
IGURE 2-

		FIC	FIGURE 2-13		Page 1	13 of 20	
TCGCGGCCGC	CCTGGCCGGC	TCGCGGCCGC CCTGGCCGGC GCCCATAGCC CCGTCATGTC	CCGTCATGTC	TGCACCCACC GTCGCCGAAT	GTCGCCGAAT	9420	
GCCTGATTGT	CTTGACCGCC	GCCTGATTGT CTTGACCGCC CGTCACGGCC	CCGTTGCGCG	CACGATCTTC GAACGACTTC	GAACGACTTC	9480	
GCAGCGAAAT	CGGCTTGAGC	GCAGCGAAAT CGGCTTGAGC GTGTCATCTT TCACCGCCGA GCATGCCGCT GCCACGCAAC	TCACCGCCGA	GCATGCCGCT	GCCACGCAAC	9540	
GAGCCTTTCT	GCGATACGGC	GAGCCTTTCT GCGATACGGC AAGGGGCGCC ACCGCGCGC TCTCAACTTC GGAGACTGTA	ACCGCGCGGC	TCTCAACTTC	GGAGACTGTA	0096	
TGACGTACGC	GACCGCCCAG	TGACGTACGC GACCGCCCAG CTGGGCCACC AACCACTGCT GGCCGTCGGC	AACCACTGCT	GGCCGTCGGC	AACGACTTCC	0996	
CGCAAACCGA	CCTTGAGTTC	CGCAAACCGA CCTTGAGTTC CGCGGCGTCG	TCGGCTACTG	GCCAGGCGTC	GCGTAACCGT	9720	35/
ATGCGCGGTG	ATCGCTGTTT	ATGCGCGGTG ATCGCTGTTT GTAATGAGTT CAGCGACACG AAGAATAAAA TATGGGTAGC	CAGCGACACG	AAGAATAAAA	TATGGGTAGC	9780	63
CGAAATCACT	AAGCTACAGT	CGAAATCACT AAGCTACAGT GCTGGTGCAC GCCATGAAAG ACCGTCAATG ACAAGGAGGA	GCCATGAAAG	ACCGTCAATG	ACAAGGAGGA	9840	
CGGCCGAAAT	CGGCCGAAAT GCCCAAGGAC CGA	CGACTGCCGG	ACTTGACGCC CACAGGAGCG		TACGCACCGG	0066	
CCAACAGCGG	CCAACAGCGG CATGACCATG GCA	GCAAGGCAGG	ACGGCCCTCG	ATGACCGGCA	AGCGCGTTGA	0966	
GCGGGTGCAC	GCGGGTGCAC GCAATCAATT GGA	GGAACCGGTT	GCTCGATGCT	ACCGGTT GCTCGATGCT AAAGATTTGC AGGTCTGGGA	AGGTCTGGGA	10020	
ACGTTTGACC	GGTAACTTTT	ACGITIGACC GGIAACTITI GGIIGCCGGA AAAGAITCCG CICTCCAACG ACCIGGCAIC	AAAGATTCCG	CTCTCCAACG	ACCTGGCATC	10080	
TTGGCAAACG	TTGGCAAACG TTGAGTTCCA CCG		GACGACGATC	AGCAGCA GACGACGATC CGGGTGTTCA CCGGCTTGAC	CCGGCTTGAC	10140	

FIGURE 2-14	Page	Page 14 of 20
CCTGCTCGAC ACCGCGCAGG CGACGGTGGG AGCAGTGGCC ATGATCGACG ACGCGGTCAC	ACGCGGTCAC	10200
CCCCCACGAA GAGGCGGTCC TGACCAACAT GGCGTTCATG GAGTCAGTGC	ACGCCAAGAG	10260
CTACAGCTCG ATCTTCTCGA CCCTGTGCTC GACCAAGCAG ATCGACGATG	CCTTCGACTG	10320
GTCGGAACAG AACCCTTACC TGCAGCGAAA AGCGCAGATC ATCGTCGACT ACTACCGCGG	ACTACCGCGG	10380
TGACGACGCG CTCAAGCGCA AAGCATCGTC GGTAATGCTG GAGTCCTTCC	TGTTCTACTC	10440
CGGCTTCTAC CTGCCCATGT ACTGGTCGTC GCGGGGTAAG CTCACCAACA	CCGCCGATCT	10500
GATCCGGCTG ATCATCCGAG ATGAAGCCGT CCACGGCTAC TACATCGGCT ACAAATGTCA	ACAAATGTCA	10560
ACGAGGTTTG GCCGACCTGA CCGACGCCGA GCGGGCCGAC CACCGCGAAT ACACCTGCGA	ACACCTGCGA	10620
GCTGCTGCAC ACGCTCTACG CGAACGAGAT CGACTATGCG CACGACTTGT	ACGACGAGTT	10680
GGGCTGGACC GACGACGTTT TGCCCTACAT GCGTTACAAC GCCAACAAGG CGCTAGCCAA	CGCTAGCCAA	10740
CCTGGGATAC CAGCCTGCAT TCGATCGTGA CACCTGCCAG GTGAACCCGG	CCGTGCGCGC	10800
AGCTCTCGAC CCCGGTGCAG GGGAGAACCA CGACTTTTTC TCCGGCTCCG	GAAGCTCATA	10860
CGTAATGGGC ACCCACCAAC CCACCACGGA CACCGACTGG GACTTCTAAC CGCCCAGGGC	CGCCCAGCGC	10920

15	
<b>XE 2-</b>	
FIGUR	

7	80	40	00	09	20	08	40	00	09	20	80	40	0.0
rage 15 or 20	10980	11040	11100	11160	11220	11280	11340	11400	11460	11520	11580	11640	11700
rage	TTGAGTCTGG	ATATCGCGGC	AATTTCGACG	TCGGGTGTGG	GGCCAGGCGT	AGTTACTACC	CTCAGCGCGG	GCCACATACC	CAGAACGCCG	TATCGCAAGC	GTCTTCCCCA	CCGAATGCCG	ATTTTCTTCT
	ATCTGCTAGC	TATGTTTGTC GTCGACTCAG ATATCGCGGC	CTCCCGGAGG	GCTCTGTTGT	CACCGATACC	CAGCCTGCCC	CGACAAGTTC	TATCACCTCG GCCACATACC	CAAGGTCTAC CAGAACGCCG	GGACCAGGCC	GCTGCCAGTC	ATCGATAGCG CCGAATGCCG	CGACGGGGTG
FIGURE 2-13	CGCGACACCG GGCCCGATCG ATCTGCTAGC TTGAGTCTGG	TATGTTTGTC	CTGCAAACTA	CTGTCGTTTT	TACTGCGAGG AGTTGAAAGG CACCGATACC GGCCAGGCGT	CCGGCCTACA ACATCAACAT CAGCCTGCCC AGTTACTACC	AATTACATCG CCCAGACGCG CGACAAGTTC CTCAGCGCGG	GAAGCCCCT ACGAATTGAA	CCGTGGTGCT	ACGTACAAGG CCTTCGATTG GGACCAGGCC TATCGCAAGC	TGGCAGGCTG ACACCGATCC GCTGCCAGTC GTCTTCCCCA	AAGCAGACCG GACAACAGGT	CAGTCACGAA CGACGGGGTG
T L	CGCGACACCG	CGCGATGCCC	GCCGGCGGTG	CTGGTCACGG	TACTGCGAGG	CCGGCCTACA	AATTACATCG	GAAGCCCCCT	GGTACGCAGG	ACGTACAAGG	TGGCAGGCTG	AAGCAGACCG	CAGAACTTCG
	GTCGGGGGCG TCGAGCACCA	TCAGGCATCG TCGTCAGCAG	AATCCAATCT CCCGCCTGCG	TGCGCATCAA GATCTTCATG	CCACGGCCGC GCCCAAGACC	GCCAGATTCA AATGTCCGAC	CCGACCAGAA GTCGCTGGAA	CCACATCGTC CACTCCACGC	AGTCCGCGAT ACCGCCGCGT	GCGGCACGCA CCCAACGACC	CAATCACCTA TGACACGCTG	TTGTGCAAGG TGAACTGAGC	GCTTGGACCC GGTGAATTAT
	GTCGGGGGCG	TCAGGCATCG	AATCCAATCT	TGCGCATCAA	CCACGGCCGC	GCCAGATTCA	CCGACCAGAA	CCACATCGTC	AGTCCGCGAT	GCGGCACGCA	CAATCACCTA	TTGTGCAAGG	GCTTGGACCC

						JU/ U.	,						
16 of 20	11760	11820	11880	11940	12000	12060	12120	12180	12240	12300	12360	12420	12480
Page	CCGGCCCAAC CCAGGTATTG GTCCCACGTT	GCCTAGACTC GCGAGGACCG CGCGGTGGTC ACTGCGCGGA	TGTTCGGTGC GCCCACTGCG GTGACTCACC TGCAGCGCCG	TCTATGGTGC GTTAGAGGAT	CGAGAGGATA TGCGATCCAC	CGICAIGCIC GGGAICAACI CGAIAAICGG CGCCGGIAIC	TCGCCCCGAT GGCCTATGTT	CGACGCGGC AAGGTACGTC	ACGGCCGCAT TTGGGCGCCG GATCGGCATC	GGGGGTGTT GGCTTCTTTT	CCTGGGCCGA CGCCGAGCAA	GCGTGCTGTT GGCCATCAAC	CAAGTGGGCC AACGGAACGT CAACGGTAGG CAAGGCATTC
FIGURE 2-16	TCAACCCGGG GGAGTTGCTG CCCGAAGCAG CCGGCCCAAC	CCGCGATCGA CTCGATGCTG GCCTAGACTC GCGAGGACCG	TTTGGGGCGG CGGAAGTGAG TGTTCGGTGC GCCCACTGCG	GCATCGACAG GCCGGGAGCT CAAGAATCGT CGCTAGAGAA	TCCCTGCTAG ACAGCCTTGG TGCGGTGGTC GGCCCGCGGA	AAGCTGGGTT TCTGCAGCGT CGTCATGCTC GGGATCAACT	TTCCTAACTC CAGGTGAGGT GATCGGGCTC GCAGGACCCT	TTAGCTGGCA TTTTCGCGGG TGTCGTGGCG ATCGTCTTCG	AGAACAAACG GTGCCTCCTA CGCCTACACA ACGGCCGCAT	TATGTCGGTG TCACCCACGC CATTACCGCG TCCATCGCTT	TTCGTCTCGA CGCTGTTGCG AGTGGCCTTC CCCGACAAGG	CTGTTCAGTG TGAAGACGCT GACGTTTCTC GGCTTTATCG	CTCTTCGGCA ACCGGGCGAT CAAGTGGGCC AACGGAACGT

						39/6	3						
17 of 20	12540	12600	12660	12720	12780	12840	12900	12960	13020	13080	13140	13200	13260
Page	CGTGAACAAC	CGTCGCCGAA	ATTGGCAAGG GCACGTTCTC GAGTATGGCG CTGGCCACGA TTGTCGCGTT GTACGCATTC	ACCGGTTTCG AATCGATCGC GAACGCCGCC GAAGAAATGG ACGCGCCGGA CCGGAACCTG	CACCCTAACG	GAAACTGGCC	GCGGCCATCG GAAACGCTAC CTTCCGAACG ATCATCGTCG TCGGAGCCCT GATATCGATG	GCACCGCGGC TTTGGACCGC GTTAGCGGAC	GATGGTCTCC	CGACAACCTG	CACCTGACCG GCCTGGCGGT GATCGCCCGA TTCGTCCAGT TCATCATCGT GCCGATCGCT	CTGTGCGGCG AAATGCGTTC	AGTGTCCTAC
	CCACCCAGCA	CGTTGCTTGG	TTGTCGCGTT	ACGCGCCGGA	TCTACTTGCT	GCGACACCGT	TCGGAGCCCT	TTTGGACCGC	ACGACGTGCC	CGCTGCGGTT	TCATCATCGT	CTGTGCGGCG	TTGGGCTGGC
FIGURE 2-17	TGGATCATCA CCACCCAGCA	ACCCCGTACT	CTGGCCACGA	GAAGAAATGG	GTTGGCGCGA	GCCGCGTCGG	ATCATCGTCG	GCACCGCGGC	AAGAACCAAT	TTCCCGTTGG	TTCGTCCAGT	GAACATGCTG	GTGGTCTCGG
FIC	CGGCGGGCTG	ATACAGCGCG	GAGTATGGCG	GAACGCCGCC	GATCTTCTCG	GAACAAGATC	CTTCCGAACG	CTCGTTCGGT	CTTGTCACGC	GGCGCTCGCA	GATCGCCCGA	TCAGGCAGTA	
	GCGCTCTCGG CATTCATTGT CGGCGGGCTG	TACGCAACGG CGTGGTCGGC ATACAGCGCG	GCACGTTCTC	AATCGATCGC	CCGAGAGCTA TACCGATCGC GATCTTCTCG	GTAGCGATGC TGCTCGGATC GAACAAGATC	GAAACGCTAC	Tregeratea atgregege crestreger	AGCGGGGTTC TGCCGACACG CTTGTCACGC	Tregeratta eggegreget ggegereger treeegtigg	GCCTGGCGGT	CTCATCGCAT TGGCGAGGTC TCAGGCAGTA	ACCGACAAGG TGTTACCGCT TGTTGCGATC
	GCGCTCTCGG	TACGCAACGG	ATTGGCAAGG	ACCGGTTTCG	CCGAGAGCTA	GTAGCGATGC	GCGGCCATCG	TTCGGCATCA	AGCGGGGTTC	TTCGCAATTA	CACCTGACCG	CTCATCGCAT	ACCGACAAGG

13320	13380	13440	13500	13560

) ) ) }	13440	13500
)	A	A

13620

13680

13740

CAGGCGACCG

CCCAGCTCTC

13860

13920

13800

CGTTTGAGGT

40/63

GACTACCGCT GCATCTTTCT AGTGCGGGT GGTCCGAACT ACTTCTCGAT TGCTTTGATC

TCATCGTGGT ACCGGCGATG

GTGATCACGT

FIGURE 2-18

20

ο£

18

Page

GCTTATCTGC ACTACTACCG AATCATTCGC

CGGGTTGGCG ATCGGCCGAG CACTCGCTAG ATTCCGTTGG CGCTGAGCTC GAACGGGAG

ACACAACGGC GAGCGATGGC GGGAATAGCC TGGTCGGTGC GGGCAAGATT TCAACCTGC

TTCCCGGATC GGCGCGCGC GCAAGCGTCT GCAACGCCGA GGGACTGTAG GCACGTAGTG

CGCTGATAAA GCCGTCGTGC ATGCTCGAGC GCATCGACGA CCATGGCAGC AGCAGTAGGT

SUBSTITUTE SHEET (RULE 26)

TAGCAGCACC GAAGAGAGCG TGAACGACAG CGGTTTCTGC GGAGCGGCAG

GTGGTGAAAT GCCAGCGCGA AGACCGCCAG GTCATAGCTG TGGTCGTGGC AAAGCGCTTC CCCACCCGGG TGGCCTCGGC GATCGCTTTG TAGGCGGCAG CGATGATCAG

CGGATGTGTT GGTGGCGTCG ATCACTTGGG TGCGTGCTCG CGTCGATTGC CCGCGCGAT GTTGGCCACC GAGGTGGGAT CTAGATCGCT GATCGTCACC GTCGCTGTCG GGTGTAGCTC GAGGATTTTC GCTGAGAGCT TGCCATGGCC CGCACCAAGT TCCAGGATTC

GCGGGTTGGG AATGTCAGAA ACAAGTTTCA GGGCTATCCG GGCGTACTTC TCGTGCAGGT

13980 14040

	FI	FIGURE 2-19		Page	Page 19 of 20
TGGTCAGGGT GCCCACCCGG		TCGAGCACCC CGATGATCTT	CTGTTTGACC TCATCGGGCA	TCATCGGGCA	14100
CATCGTCGCG GTCGAGGTAC	TCCAGTGCGT	CGGTCTGGAA	TCGACGATCC	AGCCAAGACG	14160
CGTCGGGGCC ACCCCGTGGC	: ATCGTGGCGA	TCGCCTGCTC	TCGCCTGCTC GCGGATGTTC GCCTCACCCA	GCCTCACCCA	14220
TGGCAGCTCT TCCCCTCTCG	ACGICCCGIG	TTCGCAATGC	TTCGCAATGC TATGAGACCG CTGACCGGGC	CTGACCGGGC	14280
TCCCCAGCCC GCCGGTCGCG	GGGCTTAGC	TACGTAGCAG AGGGGCCGTC		ACTTCGAGGG	14340
CTGCCGCCAC TCGGTGATCT	TGCGGCCCAA	TGAATCGGCC GCGTTCGAGG		CTGCCCGTCC	14400
CACGGCTTTG GTTCACGGTG		AGCCGGTGCC	AAGATCGCAC AGCCGGTGCC GGAAAAGTCC GCGGCACCGA	GCGGCACCGA	14460
TGTCGGTCAG CAAGACGTTG		CCGAGATCAC	AAGAGAAACC CCGAGATCAC CGCCCATGGG ATCGTCATCA	ATCGTCATCA	14520
ACACCCCAGG CAGCGTCGAC		ACCCGCCCCA CGAACCAGCA CTGAAGTAGG		TATTCACGCC	14580
ACGCGAAAGG CGGCTTGAAC		ACGTGTCGAG	ATGCACACGG ACGTGTCGAG CGTCATCGCG AAGAAATCGC	AAGAAATCGC	14640
CCAGCGCC CACCGGCCGC		CAGCGACCCG	AAGACCGGAT CAGCGACCCG ACCGGCCGCC TTGTCGGCCA	TTGTCGGCCA	14700
CGATTACCAT GGCGCGCGC	: ACCAGCTGGA	TTCGATGCTG GGCCGTTGGG		TGAGGTGGCG	14760
CACGCTGGCC CCCCGGACAG		GTCGACGATC GGTGACATTG	GTGAGCGTAC GCGGCAGAGA	GCGCCAGAGA	14820

					•	L, 00	
Page 20 of 20	14880	14940	15000	15060	15120	15180	15239
Page	ATAGCAATGC	GGGCGATGTG	CGAGTCACGC	CGGATGCCGC	GAGGTTGCCA	CCCCACTCGC	GGCGAATTC
	TGATCGGTAA	CTATAGCGCG GGGCGATGTG	GCCGTTGGCT	GGGCGGTGAT	CGCGACCGGC	TTTTGCCAGC	CGTTGCGCAC
FIGURE 2-20	TACGCGATTG CTTGGACAAC TGATCGGTAA ATAGCAATGC	TGATGTATCT TGCTAGTATC	ACAGGCGCAT CACCGGTCAA GCCGTTGGCT CGAGTCACGC	SCATCAACAG CGCGCCCGAC GGGCGGTGAT CGGATGCCGC	GGCCCGCCGA CCAGAGCCTT CGCGACCGGC GAGGTTGCCA	TGCTGCTAAC GAGCCTGTAG TTTTGCCAGC CCCCACTCGC	SGCTCAGCGA CGGCTCATGT CGTTGCGCAC GGCGAATTC
FIG	TACGCGATTG	TGATGTATCT	ACAGGCGCAT	GCATCAACAG	GGCCCGCCGA	TGCTGCTAAC	GGCTCAGCGA
	CCGCTGATGT CCATAGCCAA	AAACTGGCAT ATATTGGCTA	CTCTGCTGCC TTGGCGGCCG	TGGCGAGGCA CCACGATCAG	ATCCTGACCG CCTCGATTCG	CCATGGTCGT CGAAGCAACT	CAGGTTTTCA
	CCGCTGATGT	AAACTGGCAT	CTCTGCTGCC	TGGCGAGGCA	ATCCTGACCG	CCATGGTCGT	GCTTTGTCTG CAGGTTTTCA

						43/6	3						
1 of 16	09	120	180	240	300	360	420	480	540	009	099	720	780
Page 1	GAATTCACTT AGCTAACACC AGTTCTAGCA GCTGTCGGCG CGACTTCTTG TCAGTGCCCG	CTTGTCGAGG TCTTTGATGC	GCTGGATGCC GAGCTGGACC GCTTGGACGA GGTGTCTTTT GAGGTGTTGA CCACCCCGGA	CTGGAATG CTTGGTGCGC CGGCTACCGG CGGTCGGGCA	SAACTGGGCG GCACGCTGTG	CAAGCCCGAC GCCGCCCTAC GCATCGCCGA	CGCCGCCGAT CTCGGACCTC GTCCGAGCAC TCACCGGCGA ACCGCTAGCC CCACAGTTTG	ACCGCCACCG CCACCGCCCA ACGCCAGGGC CTGATCGGCG AAGGCGCACA TCAAAGTGAT	GTGTCCAAAC CCGCCAGGCC	GCCGAAGCCC GACCTGGCCG CAAACCGCTC AAATATCGTC CCGACGAGCT GGCCCGCTAC	GCCCAGCGGG TCATGGACTG GCTACACCCC GACGGCGACC TCACCGACAC CGAACGCGCC	CAATACGACG GCATGTCACG GCTAAGTGGC	TAGCCAAACT GGCCGCCCCC
FIGURE 3-1	A GCTGTCGGCG	CCGGGAGGAG	A GGTGTCTTTT	3 CTTGGTGCGC	CAGCGAGGAA GAACTGGGCG		TCACCGGCGA	CTGATCGGCG	CGCGGTGGAT	AAATATCGTC	GACGCCGACC		TACCTGACCC CCCAAGCGCG GGCCACCTTT GAAGCCGTGC
<b>\( \)</b>	AGTTCTAGC	TAGCGAATA	GCTTGGACG	GTCTGGAAT	ACACCCAAG	TACGCATCA	GTCCGAGCA	ACGCCAGGG	ACCTGCCCG	CAAACCGCT	GCTACACCC	GAGCAACCAC	GGCCACCTT
	AGCTAACACC	ACGTTATGAT TCGAACATGT TAGCGAATAG	GAGCTGGACC	ACGGCTGCGG TCTCTGGAAC GT	CACGTTGATC AACCAACTCG ACACCCAAGC	CTGCGCGCTG GCCAACCGGT TACGCATCAC	CTCGGACCTC	CCACCGCCCA	TCGCGCCCTT TTTCGGCCCA ACCTGCCCGC	GACCTGGCCG	TCATGGACTG	CGCAAACGCG GCATCACCCT GAGCAACCAG	CCCAAGCGCG
	GAATTCACTT	ACGTTATGAT	GCTGGATGCC	ACGGCTGCGG	CACGTTGATC	CTGCGCGCTG	CGCCGCCGAT	ACCGCCACCG	TCGCGCCCTT	GCCGAAGCCC	GCCCAGCGGG	CGCAAACGCG	TACCTGACCC

of 16

FIGURE 3-2

1500

1560

1440

TGCTGTCTAC CTCCGAATCT CAGAAGACCG CTCCGGCGAA CAGCTCGGCG TGGCCCGCCA

ACGCGAGGAC TGCCTAAAGC TGTGCGGGCA GCGAAAATGG GTGCCCGTCG AGTACCTCGA

7	
Page	TGCGGCCGCC
	CGA

ACCA	GGCGCGACCA ACCCCGACGA	CCACACCCCG	CCACACCCCG GTCATCGACA CCACCCCGA TGCGGCCGCC	CCACCCCGA	TGCGGCCGCC	840
Ø	ATCGACCGCG ACACCCGCAG	CCAAGCCCAA	CCAAGCCCAA CGCAACCACG ACGGGCTGCT GGCCGGGCTG	ACGGGCTGCT	GGCCGGGCTG	006
Ţ	CGCGCCTGA TCGCCTCCGG	GGAACTGGGC		CAACACAACG GTCTTCCCGT	CTCGATCGTG	096
$\circ$	GTCACCACCA CCCTGACCGA	CCTGCAAACC	CCTGCAAACC GGCGCCGGCA AGGGCTTCAC CGGCGGCGGC	AGGGCTTCAC	ລອອລອອລອອລ	1020
O	ACCCTGCTAC CCATGGCCGA	TGTGATCCGC	TGTGATCCGC ATGACCAGCC ACGCCCACCA CTACTCCCCC	ACGCCCACCA	CTACTCCCC	1080
$\cdot$	GCAAGCGGGA GGTACCCCCA	GGCGATCTTC	GGCGATCTTC GACCACGGCA CACCCCTGGC GCTGTATCAC	CACCCCTGGC	GCTGTATCAC	1140
	ACCAAACGCC TAGCCTCCCC	GGCCCAGCGG	ATCATGCTGT	TCGCCAACGA CCGCGGCTGC	CCGCGGCIGC	1200
$\cdot$	ACCAAACCCG GCTGTGACGC	ACCGGCCTAC	ACCGGCCTAC CACAGCCAAG CCCACCACGT CACCGGCTGG	CCCACCACGT	CACCGGCTGG	1260
$\cdot$	ACCAGCACCG GACGCACCGA	CATCACCGAC	CATCACCGAC CTCACCCTGG CCTGCGACCC CGACAACCGA	CCTGCGACCC	CGACAACCGA	1320
7	CTCGCCGAAA AAGGCTGGAC	CACCCGCAAA	CACCCGCAAA AACACCCACG GCCACACCGA ATGGCTACCA	GCCACACCGA	ATGGCTACCA	1380

••	
8	
<b>ن</b>	
<del></del> 3	
3-3	
3-3	
, 3-3	
3-3	
E 3-3	
<b>LE 3-3</b>	
<b>RE 3-3</b>	
<b>RE 3-3</b>	
JRE	
IGURE 3-3	
GURE	
JRE	

	FIGURE 3-3	Page 3	Page 3 of 16
CAACGACGTC AGCGCATCAA CCGGCAAGCG	CCGGCAAGCG CCGCCCGCC TACGAGCAGA TGTTGGCCGA	A TGTTGGCCGA	1620
CATCACCGCC GGCAAGATCG CCGCCGTGGT	CCGCCGTGGT GGCCTGGGAC CTGGACCGGC TCCATCGCCG	TCCATCGCCG	1680
TCCCATCGAG CTGGAAGCCT TCATGTCATT	AGCCGACGAG AAGCGGCTGG CCCTGGCCAC	CCCTGGCCAC	1740
CGTCGCCGGC GACGTTGACC TGGCGACACC	TGGCGACACC CCAGGGCCGG CTAGTCGCCC	CTAGTCGCCC GCCTGAAGGG	1800
GTCGGTGGCC GCTCACGAAA CCGAGCACAA	CCGAGCACAA GAAGGCACGA CAGCGCCGCG CCGCCCGC	CCCCCCCCA	1860
GAAAGCTGAA CGCGGCCACC CCAACTGGTC	CCAACTGGTC GAAAGCCTTC GGCTACCTGC CCGGCCCCAA	CCGGCCCCAA	1920
CGGTCCCGAA CCCGACCCCC GGACAGCGCC	GGACAGCGCC GCTGGTCAAA CAGGCCTACG	CCGACATCCT	1980
CGCCGGGGCG TCCCTGGGCG ACGTGTGCCG	CCAGTGGAAC GACGCCGGGG	GTTCACCAT	2040
CACCGGCCGC CCGTGGACGA CTACAACGCT	CTACAACGCT GTCGAAATTC TTGCGCAAAC	CCCGCAACGC	2100
CGGACTACGC GCATATAAGG GTGCCCGCTA	GTGCCCGCTA CGGCCCGGTG GACCGCGACG CGATTGTCGG	; CGATTGTCGG	2160
CAAGGCCCAG TGGTCGCCGC TGGTGGACGA	TGGTGGACGA GGCGACGTTC TGGGCCGCCC AGGCCGTGCT	: AGGCCGTGCT	2220
GGACGCCCC GGCCGCGCC CCGGCCGCAA	CCGGCCGCAA AAGCGTGCGC CGCCACCTGC TGACCGGGCT	TGACCGGGCT	2280
GGCAGGCTGC GGCAAATGCG GCAACCACCT	GCAACCACCT GGCCGGCAGC TACCGCACCG ACGGCCAGGT	ACGGCCAGGT	2340

Z.
Page
3-4
FIGURE 3-6

CGTCTACGTG TGCAAGGCGT	GCCACGGGGT	GCCACGGGGT GGCCATCCTG	GCCGACAACA	TCGAACCGAT	2400
CCTGTATCAC ATCGTGGCCG	AGCGGCTGGC		CATGCCCGAC GCCGTTGACT	TGTTGCGCCG	2460
GGAGATTCAC GACGCCGCCG	AAGCCGAAAC	CATCCGCCTG	GAACTGGAAA	CCCTCTACGG	2520
GAGCTGGACA GGCTCGCCGT	CGAACGCGCC	GAAGGGCTAC	TGACCGCGCG	CCAGGTGAAG	2580
ATCAGCACCG ACATCGTCAA	CGCCAAGATA	ACGAAACTTC	CGCCAAGATA ACGAAACTTC AGGCCCGCCA ACAGGATCAG	ACAGGATCAG	2640
GAACGGCTCC GAGTGTTCGA	CGGGATACCG	TTGGGAACAC	CGGGATACCG TTGGGAACAC CGCAAGTCGC CGGGATGATA	CGGGATGATA	2700
GCCGAGCTGT CGCCGGACCG	GTTCCGCGCC	GTCCTCGACG	GTTCCGCGCC GTCCTCGACG TCCTCGCTGA AGTCGTTGTC	AGTCGTTGTC	2760
CAGCCGGTCG GCAAGAGCGG	CAGGATATTC	AATCCCGAAC	CAGGATATTC AATCCCGAAC GGGTGCAGGT GAATTGGCGA	GAATTGGCGA	2820
TGAGCCGGCA CCACAACATC	GTGATCGTCT	GTGACCACGG	GTGATCGTCT GTGACCACGG CCGCAAAGGC GATGGCCGCA	GATGGCCGCA	2880
TCGAACACGA GCGCTGCGAT	CTTGTCGCGC	CGATCATTTG	GGTCGACGAG ACCCAGGGCT	ACCCAGGGCT	2940
GGTTACCGCA GGCGCCAGCG	GTGGCAACAT	TACTCGACGA	TACTCGACGA CGACAACCAG CCGCGAGCCG	CCGCGAGCCG	3000
TTATTGGCTT GCCGCCCAAC	GAGTCTCGCC	TACGACCTGA AATGCGCCGC	AATGCGCCGC	GACGGGTGGG	3060
TGCGGCTGCA CTGGGAATTC	GCCTGCCTGA	GGTACGGCGC	CGCCGGCGTG	CGCACGTGCG	3120

16

Page
P
Ċ
ų
H
10
FIGUR
国

	3660 3720 3780 3840 3900	GCCGGCGGGT CGCCGCCGAC GCGACGATGC TCGGCCACGG GTCGCCACAC CGTCTAGCGC CGATCCTACC GCGTCACGCG CACGAGCATG TCGCGCCGGT CCTGGATGCT GCCGGGTCGT GCCTGGCGTC AGCTCGACGA CGCCGATCCT CGCAAATGGG CGGCACTGG CTCTGAGGGT AGAGACGTGC CAGGAGGCGA	CGCCGCCGAC GCGACGATGC CGTCTAGCGC CGATCCTACC TCGCGCCGGT CCTGGATGCT AGCTCGACGA CGCCGATCCT CTCTGAGGGT AGAGACGTGC	GCCGGCGGGT GTCGCCACAC CACGAGCATG GCCTGGCGTC	GGTGCCCGAT GACCGCCGGC AGGACCGGGC ACCCGCGACA CCGTGTCGTG GTGGTCGGTG GGCCGATGGC CGGCACACCG	GGTGCCCGAT AGGACCGGGC CCGTGTCGTG GGCCGATGGC
	3600	CCGGCTGGCC GAGCTGCTGC GCCCCGTGCG GCGGTCAGGC	GAGCTGCTGC (	CCGGCCGGGT	CTGACTGACG AGCAGCGCAC	CTGACTGACG
_	3540	CCCTCGCCGC AGCGCCGCCC	CTCATCCGCA	CGCGGACTAC	GATCTCGTTG CCGCGAACAT	GATCTCGTTG
	3480	AGTTCATCGC CGCGAAAACC	GACGATCCCG	CCGCGCATCC	GTAGCCGCAC TGTCTCGTTC	GTAGCCGCAC
	3420	GGAGTCGCAC GTGTCGACCA TCTACCATCA TCGCGGCCGC	GTGTCGACCA	GGAGTCGCAC	CGGGGTTTTG GTGTCCCCAA	CGGGGTTTTG
	3360	CGTGGTAGCT GCTAGACTCC GACGTAGCCG GCTTCGACTC	GCTAGACTCC	CGTGGTAGCT	AACTACGCGC CACCCCAACG	AACTACGCGC
	3300	TGGCTGTGGC GCACGCTCTG CGAAAGCGAC ACGCCGAACA	GCACGCTCTG	TGGCTGTGGC	ACGCGGTGGT CGTCGCCATG	ACGCGGTGGT
	3240	GGCAACCCCG ACTACGCACC GGGTTTTGCG GTGCAGTCGG	ACTACGCACC	GGCAACCCCG	GGCTACTGAC CGGACTGGCC	GGCTACTGAC
	3180	CGCAACGGCG ACCTGCAAAC ACTGTGCGAG AACGTTCCGC	ACCTGCAAAC	CGCAACGGCG	AGCAGCGCC CGTGCGGGTT	AGCAGCGGCC

						48/	63						
6 of 16	3960	4020	4080	4140	4200	4260	4320	4380	4440	4500	4560	4620	4680
Page (	GCCCGGCATC GCCCGCGAGA	CCGACATCCC	TGCTCATCGA	ACCGGCTAGC	CCTGGGACGG	CAGAGCTGCG	GAAAATGCGA	CATTCGCCGC	GGACGCTGGA	AGATATGCCG	CCCGCGTTCT	TACTAGGCAC	AATCTGTGTT
	GCCCGGCATC	GGCGTGATGG	GACGGGCACG	CGACGGTGGC AAACCGACGC AAGTTCATCG CGGCCAAGCC CGAATCGCCT ACCGGCTAGC	CGAACGTTAC CAGGACAAGC TGCTGCACGT GGCCGGGATC GGCTGGCACT CCTGGGACGG	AGCCAAACGT GCAGTGCTGG	GCCGACGTCC	ATCGGCGTCC GGCGTGGCCG GCGTGCTCGA CCTGGCCGCC GCACTGGTAC CATTCGCCGC	GACGCTAGCC GACCTCGACA GCGACCCGCA CTTGCTCAAC GTCGCGAATG GGACGCTGGA	CGCATCACAA	CGGTGCCTAC CAGTCCGACA CCGAATCGCC TCTCTGGCAA GCGTTCTTGA CCCGCGTTCT	GCCCGATGAA GGTGTGCGCG GGTTCGTGCA ACGCCTGGCC GGCGTCGGCC TACTAGGCAC	CGTCCGCGAA CATGTCCTGG CGATTCTTAT CGGTGTAGGT GCCAACGGAA AATCTGTGTT
FIGURE 3-6	TGGCGCAGGC GTCACGTGAC GTATCTGCGG CCGCCGACTG	GGGCGGGGGT	ACGCCCCCTG GATCGACCGG GACGGGCACG	CGGCCAAGCC	GGCCGGGATC	AGCCAAACGT	CCAAGCGCTC TCAGACAGCC TCAACGACAA GGAATTACGC GCCGACGTCC	CCIGGCCGCC	CTTGCTCAAC	GCCCGCTGAC	TCTCTGGCAA	ACGCCTGGCC	CGGTGTAGGT
F	GTATCTGCGG	TACATCCCGC	ACGCCCCCTG	AAGTTCATCG	TGCTGCACGT	ACCGCGGCGA	TCAACGACAA	GCGTGCTCGA	GCGACCCGCA	GGCCCCACGC	CCGAATCGCC	GGTTCGTGCA	CGATTCTTAT
	GTCACGTGAC	TCGTCCGACG GCGCGGCGTG	CTACGGCACC GACTATCCCG	AAACCGACGC	CAGGACAAGC	CAGACGCTGG GCAGCCGACG	TCAGACAGCC	GGCGTGGCCG	GACCTCGACA	CCTGCACACG CTCAAATTGC GGCCCCACGC	CAGTCCGACA	GGTGTGCGCG	CATGTCCTGG
	TGGCGCAGGC	TCGTCCGACG	CTACGGCACC	CGACGGTGGC	CGAACGTTAC	CAGACGCTGG	CCAAGCGCTC	ATCGGCGTCC	GACGCTAGCC	CCTGCACACG	CGGTGCCTAC	GCCCGATGAA	CGTCCGCGAA

		FI	FIGURE 3-7		Page 7	of 16	
CGACAAGGCG ATTCGCTATG CC	ATTCGCTATG	CCCTTGGCGA	TTATGCCTGC	CTTGGCGA TTATGCCTGC ACCGCTGAGC	CTGACCTTTT	4740	
CATGCACCGG GAAAACGCTC AC	GAAAACGCTC	ACCCAACAGG	CCAACAGG CGAAATGGAC CTCCGCGGCG		TGCGATGGGT	4800	
AGCGGTATCC (	GAGAGCGAAA	AAGATCGCCG	GCTGGCCGAA	AGCGGTATCC GAGAGCGAAA AAGATCGCCG GCTGGCCGAA TCAACGATAA AACGGCTGAC	AACGGCTGAC	4860	
TGGCGCGAC GCCATCCGCG CC	GCCATCCGCG	CGAAAGAT	GCGCCAAGAC	TTCGTGGAAT '	TCGAGTGGTG	4920	
CCGTTTGAAG TAGTGATTCC TG	TAGTGATTCC	TGCCGACGAG	CAGGACCGGG	AACTGGACGC	ACGGTTGCAG	4980	•
TTGGAGGCCG 7	ACAGCATCCT	GTCCTGGGCG	GTGGCCGGAT	TTGGAGGCCG ACAGCATCCT GTCCTGGGCG GTGGCCGGAT GGAGCGACTA TCAGCGAATC	<b>PCAGCGAATC</b>	5040	49/
GGACTATCCC ;	AGCCGGACGC	GGTGCTCGCG	GCAACGTCGA	GGACTATCCC AGCCGGACGC GGTGCTCGCG GCAACGTCGA ATTACCGCGA GGACTCCGAC	GACTCCGAC	5100	63
ACGATAAAGA GGTTCATCGA CGACGAATGC	GGTTCATCGA		GTCACCAGCT	CGCCGGTGCT GAAAGCCACT	SAAAGCCACT	5160	
ACTACGCATC TGTTCGAGGC GTGGCAAAGG	rgttcgaggc		TGGCGGGTGC	AAGAAGGCGT 1	ACCCGAAATC	5220	
TCGCGCAAAG (	CGTTCGGCCA	GTCGCTCGAC	ACCCACGGAT	TCGCGCAAAG CGTTCGGCCA GTCGCTCGAC ACCCACGGAT ACCCGGTCAC TGACAAGGCC	rgacaaggcc	5280	
CGTGATGGTC GTTGGCGGGC CGGAATAGCG	3TTGGCGGGC		GTGAGAGGGG	GTGAGAGGGG CCGATGATTT CGATGATTAG	CGATGATTAG	5340	
CACACCTAAC GTGACGCATG TGACGCATTT	STGACGCATG		CCAGGTTCGC	CTACGCGCGC	GCACGTATGG	5400	
CGGTTATACC GCGCAAACGT CACATGCGTC	GCCAAACGT		ACGGCCTGCC GTGCCGTTCT	GTGCCGTTCT G	GCCCAGGATG	5460	

						50/6	3						
8 of 16	5520	5580	5640	5700	5760	5820	5880	5940	0009	0909	6120	6180	6240
FIGURE 3-8	CGGTACCTAC CTGGCCGTTC ACGGCCGCCA CCGGGCGGAC TGTACCGCCA AACCAGCAAA	CACCGGCGGT GCCGCATGAC CGCTGTCGCG ATCACCCCGG CATCCGGCGG TCGGCACAGC	GTCCGATTCG CCTACGACTC TGCGATCGTG TCGTTGATCA AGTCCTCGAT CCCCGCCTAT	GCCCGCTCCT GGTCCGCGCA CACCCGCTGC TGGTTCATCG ACGCTGACTG GACCCCACTG	CTGGCCGCCG AGCTGCGCTA CCACGGCCAC ACCGTCACCG GACCCGCCGA CCCGGCGCAA	CAGCAGTGCA CCGACTGGGC CAAAGCGTTG TTCCGGGCGG TCGGACCCCA GCGGACACCC	GCCGTGTACA GGGCTTTATC CAAAGTGCTG CACCCCGACG CCCCAACCGG ATGCCCGATA	CTGCAACAGC AGCTCAATGC CGCCAGAACC GCACTTACCA ACCCTGCTTG AAAGGACACA	AGCCATGGCT GAAACCCCCG ACCACGCCGA ACTGCGGCGA CGAATCGCCG ACATGGCTTT	CAACGCCGAT GTCGGTATGG CGACCTGCAA ACGCTGTGGT GACGCCGTGC CGTACATCAT	CCTGCCGAAC CTGCAGACCG GCGAACCCGT CATGGGTGTC GCCGACAACA AATGGAAGCG	CGCGAACTGT CCCGTCGACG TCGGTAAGCC GTGCCCGTTC CTAATCGCCG AGGGTGTCGC	CGACAGTACC GACGACACCA TAGAGGTCGA CCAGTGACCC CGATCAACCG GCCCCTGACC

	FIC	FIGURE 3-9		Page 9	of 16
AACGACGAAC GACAACTGAT	GCACGAGCTG	GCAGTCCAGG	TTGTCTGCTC GCAGACGGGT	GCAGACGGGT	6300
TGCTCACCCG ATGCGGCGGT	CGAAGCACTC	GAATCCTTCG	CGAAAGACGG	AACACTTATC	6360
CTCCGCGGCG ACACCGAGAA	CGCCTACCTC	CGCCTACCTC GAAGCCGGAG GCAATGTTCT TGTCCATGCC	GCAATGTTCT	TGTCCATGCC	6420
GATCGTGACT GGCTTGCCTT	CCACGCGTCG	CCACGCGTCG TATCCCGGCA ACGACCCGCT GCGAGACGCC	ACGACCCGCT	GCGAGACGCC	6480
CGACCTATCG AGCAGGACGA	CGACCAGGGG	GCGGGGTCGC	CATCGTGACC	AGGCCCAGCC	6540
CGGACACCGC CACGGTGCCG	GCGCGCATGC	ACGCTCATTA	CCTAGACTAA	AAATTGATGG	0099
GAGGACCGAT GCCAAGACCA	CCGAAACCGG	CCGAAACCGG CCCGGCTCAA ACTGGTTGAG GGCCGCTCCC	ACTGGTTGAG	GGCCGCTCCC	0999
CCGGCCGCGA TTCCGGCGGC	CGGAAAGTCC	CGGAAAGTCC CCGAGTCGCC GAAGTTTATC CGTCAGGCAC	GAAGTTTATC	CGTCAGGCAC	6720
CGGATGCCCC GGACTGGCTC (	GACGCCGAGG	CGCTGGCCGA	ATGGCGGCGC	GTCGCACCGA	6780
CTTTGGAGCG GCTTGACCTG (	CTCAAACCTG	AGGATCGGGC GCTCCTGTCC GCGTACTGCG	GCTCCTGTCC	GCGTACTGCG	6840
AGACCTGGTC CGTCTACGTC (	GCGGCGGTTC	GCGGCGGTTC AGCGGGTCCG CGCCGAAGGC CTCACAATTA	CGCCGAAGGC	CTCACAATTA	0069
CCTCACCGAA ATCCGGTGTC (	GTGCACCGGA	GTGCACCGGA ACCCGGCGGT	GACGGTTGCG	GAGACGGCGC	0969
GCATGCATCT GCTGCGCTTG	GCCTCCGAGT	TTGGCCTGAC	ລລອລລອອລລລ	GAGCAGCGAC	7020

	FIG	FIGURE 3-10		Page	10 of 16
TGGCGGTGGC GCCGGGCGAC (	GACGGCGACG	GGCTCAACCC	GACGGCGACG GGCTCAACCC GTTTGCCCCCG GACCGGTGAT	GACCGGTGAT	7080
GACCTTTTGT GTGTGATACA	ATCGAGTTTG	GCATCTCGGC	GCATCTCGGC ATCCGCTGAC GCCGGGCAGT	GCCGGGCAGT	7140
CGCCGCGGGG CGGCTGGAAC	CCGGATAGCG	GCCGCCATGC GCCACAAGCG	GCCACAAGCG	ATTCCGCGCG	7200
TTTCTTGCGT CTGCTAGGTG	GTGGCCGAAT	TTTGAGTAGC	TTTGAGTAGC ATCCTTTTCC GCATGGCCGA	GCATGGCCGA	7260
GCTGCGGTCT GGCGAAGGCC	SAACCGTGCA	CGGCACCATC	GAACCGTGCA CGGCACCATC GTGCCCTACA ACGAGGCGAC	ACGAGGCGAC	7320
CACCGTCCGC GACTTCGACG G	SCGAGTTCCA	GCGAGTTCCA GGAAATGTTC GCTCCTGGCG	GCTCCTGGCG	CTTTTCGGCG	7380
CTCCATCGCC GAGCGCGGCC	ACAAATTGAA	ACAAATTGAA GCTGCTGGTC TCTCACGACG		CTCGAACCCG	7440
CTACCCGGTG GGCCGGGCCG 1	rtgagttgcg	TTGAGTTGCG GGAGGAGCCT	CACGGCTTGT	TCGGGGCGTT	7500
CGAGATTGCG GACACCCCGG A	ACGGCGACGA	GGCTTTGGCG	AACGTAAAAG	CTGGTGTCGT	7560
CGACTCGTTT TCGGTGGGTT 1	rccgaccgat	TCCGACCGAT CCGGGACCGT	CGCGAAGGGG	ATGTGCTGGT	7620
GCGCGTCGAA GCGGCGCTGT 1	PAGAGGTTTC	TAGAGGTTTC CCTAACCGGC	GTTCCGGCCT ATTCGGGGGC	ATTCGGGGGC	7680
ACAAATCGCC GGGGTGCGCG C	CGGAATCGCT	TACAGTCGTT	TCCCGTTCGA CAGCCGAAGC	CAGCCGAAGC	7740
CTGGCTGTCC CTACTCGATT G	GGTGAACAAT	CTATGACCGA	ATTCGACGAC ATCAAAAACC	ATCAAAAACC	7800

		FIC	FIGURE 3-11		Page	11 of 16	
TCTCTTTACC	TGAAACCCGT	GACGCGGCGA	AGCAGCTCCT	TCTCTTTACC TGAAACCCGT GACGCGCGA AGCAGCTCCT CGACAGTGTC GCCGTGTGAC	GCCGTGTGAC	7860	
CTGACCGGTG	CTGACCGGTG AGGCGGCGCA GCGTTATTCA	GCGTTATTCA		GGCGCTGACG CGCCACGCCG AGGAACTGCG	AGGAACTGCG	7920	
GGCGGAGCAG	GGCGGAGCAG CGCCGCGCG GCCGCGAAGC	GCCGCGAAGC	CGAGGAGGAG	CTGCGCCGCT	ACCGGGCCGG	7980	
TGAGCTGAGG	GTGGTGCCCG	GCGCTCCCAC	CGGCGGCGAC	TGAGCTGAGG GTGGTGCCCG GCGCTCCCAC CGGCGGCGAC GACGGCGACG CGCCGCCGGG	ದಿತ್ತದೆಂದು	8040	
CAACTCGTTG	CGGGACACCG	CGTTTCGCAC	ACTGGATTCT	CAACTCGTTG CGGGACACCG CGTTTCGCAC ACTGGATTCT TGTGTGCGAG ACGGCCTGAT	ACGGCCTGAT	8100	
GTCGTCGCGG	GTCGTCGCGG GCGGCGGAGA CCGCGGAAAC	CCGCGGAAAC	CTTGTGCCGC	ACCGGGCCGC CGCAGTCCAC	CGCAGTCCAC	8160	53/6
CTCGTGGGCG	CTCGIGGGCG CAGCGCTGGC TGGCGGCCAC	TGGCGGCCAC	CGGCAGCCGC	GACTATTTGG	GCGCGTTCGT	8220	3
CAAGCGGGTT	TCCAATCCTG	TTGCGGGGCA	CACGGTTTGG	CAAGCGGGTT TCCAATCCTG TTGCGGGGCA CACGGTTTGG ACCGACCGGG AAGCGGCCGC	AAGCGGCCGC	8280	
GTGGCGTGAG	GTGGCGTGAG GCTGCCGCGG TGGCCGCCGA GCAGCGAGCG	TGGCCGCCGA	GCAGCGAGCG	ATGGGCCTGG TGGACACCCA	TGGACACCCA	8340	
AGGCGGGTTT	CTGATCCCGG CGGCGCTGGA	CGGCGCTGGA	CCCGGCGATC	CTGCTGTCGG	GTGATGGGTC	8400	
GACGAACCCG	ATTCGGCAGG	TGGCGAGGGT	GGTGCAAACG	GACGAACCCG ATTCGGCAGG TGGCGAGGGT GGTGCAAACG ACCTCCGAGA TTTGGCGGGG	TTTGGCGGGG	8460	
CGTGACTTCC	GAAGGCGCCG	AAGCTCGTTG	GTACTCCGAA	CGTGACTTCC GAAGGCGCCG AAGCTCGTTG GTACTCCGAA GCCCAGGAGG TGTCCGACGA	TGTCCGACGA	8520	
TTCGCCAGCG	TTGGCCCAGC CGGCGGTGCC		GAACTACCGT	GAACTACCGT GGAAGCTGCT GGATTCCGTT	GGATTCCGTT	8580	

10						54/6	63						
of 16	8640	8700	8760	8820	8880	8940	0006	0906	9120	9180	9240	9300	9360
Page 12	SATCGGCA AGATTCTCGC	CTCCGGCA ACGGCGAGCC	GTGGTCGTCG GCGCGGGGTC	SCTGCCGC CAAGGTTCCA	FTTGCGGC AGGCGGAAAC	SCCGATGC TAGCCGGGAA	GCGGTGACAG CGACGAATCA	PAGAGTTG GGTCCATGGT	TCGCCGGCCG ACCGGGCAGC GCGGATTCTT	GCGTTTCGAG TTCTGAAGGT	TCTGCTT AGGGGTGCCG	CGATTCGTNN NNNNNNNNNN	NNNNNNNNN NNNNNNNNNN
FIGURE 3-12	CTCCATCGAG CTGGAGGGTG ACGCGGCGAG CTTCGTTGGC GAGATCGGCA AGATTCTCGC	GGACAGCGTT GAGCAACTGC AGACCGCGGC GTTCGTCAAC GGCTCCGGCA ACGGCGAGCC	CTCCGATCAG	AGAAGCGATT GTGGCGGCGG ATGTTTACGC GTTGCAGTCG GCGCTGCCGC CAAGGTTCCA	GGCCAGCGCC GCGTTCGCGG CGAACTTGTC CACCATCAAC ACGTTGCGGC AGGCGGAAAC	CCCATCGCT GCACGACAGT CCGCCGATGC	CATGGACAC CGTTGATTCG GCC	TCCACTGGTG CTTGGCGACT GGAAGCAATT CCTCATCGTC GACAGAGTTG GGTCCATGGT	CGGGCCGAA TCGCCGGCCG ACC	GATGTGCT GGTGCGCAAC	ATAGGGCCA GGCGTGGGCG GCCTCTGCTT	CCCTGCGG GTTGCGTTGT	NNNNNNNNN NNNNNNNNNN NNN
	CTCCATCGAG CTGGAGGGTG	GGACAGCGTT GAGCAACTGC 1	CACCGGGTTC GTCAGCGCGC TAACCGGCAC	AGAAGCGATT GTGGCGGCGG A	GGCCAGCGCC GCGTTCGCGG (	TTCGAATGGC GCGCTGAAAT TCCCATCGCT	GICTGICCIG GAAGICICCC ACAIGGACAC CGITGAITCG	TCCACTGGTG CTTGGCGACT G	GGAGTTGGTG CCTCACCTGT TCGGGCCGAA	CGCCTGGTTC AGGGTCGGAT CA	GGAGACTACC GCGTAGGTAG GA	GGCCGGCCAC GCCCGCCAAC TC	NNNNNNNNN NNNNNNNNNN N

						55/6	3						
13 of 16	9420	9480	9540	0096	0996	9720	9780	9840	0066	0966	10020	10080	10140
Page 1	NNNNNNNNN	NNNNNNNNN	CCTGTTGCCG	CATTTGTTTC	TCGTTTGAGC	AGTGTGGCGC	ACCGAGGCGA	CCGGCGACCA	TGGGTTGCGT	GCGCCGATAG	GGCCCCAAAC	TTGGCTTTCG	AGATAGCTTC
	NNINNINNINNIN NINNINNINNIN INNINNINNIN	NNNNNNNNN	NNNNNNNCC AAGCCAGAAT ATCGAGCCTG GCGGCCATGG TCGCCGCCTT	CTGCTTGGCT TTCGGCCGTT CCAGCTCGGC GATCCGGCGG CCAGCGGCGC CATTTGTTTC	CGTTGTCGAT	CCTCCTTGTC	TGATCGATGC ACCGAGGCGA	CCACCTCGC GTTGGCGCTC CTTGGCTTTC GGGCGTTCCA GCTCGGCGAT CCGGCGACCA	GTCGTCGCTG	GCCGAGGGCT	GTAGCCCATG GGCCCCAAAC	CTGCCGCTGC TTGGCTTTCG	TGTTTCTCCG
FIGURE 3-13	NNNNNNNNN	NNNNNNNNN	GCGCCCATGG	GATCCGGCGG	CTGCGGGTTG	GCTGCGCCTA	CCGGCATTGG	GGGCGTTCCA	GGATTTCTTT	CGGCGGAGAT	ATTCTGGTTT	ATGATCCT GCCCCTCGCG	CGGCGCCATT
F	NNNNNNNNN	NNNNNNNNN	ATCGAGCCTG	CCAGCTCGGC	TTTGTCGTCG	GATGCCGAGG	GGGACCAAAG	CTTGGCTTTC		CGGTAGGTGC	ACGCTCTG		CGGCGGCCAG
	NINNINNININ NINNINNINNIN NINNINNINNIN	NNNNNNNNNN NNNNNNNNNN NNNNNNNNN	AAGCCAGAAT	TTCGGCCGTT	TCCGCGAACA GGCGGATTTC TTTGTCGTCG	CGCTTGTAGG TGCCGGCGGA GATGCCGAGG	TGAGACGGCT TTGGTTCCAT GGGACCAAAG CCGGCATTGG	GTTGGCGCTC	GCGGCGCCAT TTGTTTCTCC GCGAACCGGC	TGTCGATTCG TTTGAGCCGC CG	CAGTGTCTGT TTTCGTCGAA TG	CAGAATATCG AGCCTGGCGG CC	
	NNNNNNNNN	NNNNNNNNN	NNNNNNNCC	CTGCTTGGCT	TCCGCGAACA	CGCTTGTAGG	TGAGACGGCT	CCACCCTCGC	GCGGCGCCAT	TGTCGATTCG	CAGTGTCTGT	CAGAATATCG	GCCGCTCCAG CTCCGCGATC

						56/63	3						
14 of 16	10200	10260	10320	10380	10440	10500	10560	10620	10680	10740	10800	10860	10920
Page	GC GTCCTGTTGC	AG CCCGATCATC	AG TCCGAGTTGC	CGGTTGATGG CGGCTTGTTC CGGCCCGGAG CGGCCGGCCA GGATCGCCGG GCCGCACTCG	TC GAGCTCGCCG	CT CACCTCCAAG	GGCGCGTCCG GGAACCACGC CCGGATCGTC TCGGCCACCT GGTCGCGGTC GCAGACGGCG	ACCAAGTGGT TGAAACGTGC TGGTGTGGTG	AA CCCCGCACCA	AC CACGACGATC	CGTGAGAATC GCCGCCCGCG AAGATCTTTG GACATCCCCA CATCGACGTG CGTCCTCGCC	IT GCGATCAAAC	3G CACGCGTGTT
	ACCGCCCAGC	CGCCGCCAAG	CGCCGGCG	GGATCGCC	GGAACGCT	GCGCTTCCCT	GGTCGCGG	TGAAACGT	CCACGGCCAA	CGAGGACAAC	CATCGACG	AGACGAAG	AGGTACTC
FIGURE 3-14	CGCCCGTGGG	GCTGGGCTTC	GGCTCGTCGT	CGGCCGGCCA	CTGATCCGCA GGAACGCTTC	AGCCGAACGA	TCGGCCACCT	ACCAAGTGGT	CTTCAGCAGT	GCCACAACGA	GACATCCCCA	CTGGCCATTA AGACGAAGTT	TCACCCCAGC AGGTACTCGG
FIG	ATCCATGCCC	CTCGTCTTCC	GTCGCCGTCG	CGGCCCGGAG	GTGCTCGACG	CGCGGCCTGC	CCGGATCGTC	GGTTACCGTC	GTGGAATGTT	AAGCTGCTAC	AAGATCTTTG	AACCCGCGAG	Tegecccecc
	CGGCCCATGG GCCGGAAGCT ATCCATGCCC CGCCCGTGGG	CGCGGTGTTC ACCGTCAGCG CTCGTCTTCC	CGGCTCCACG GCGGGGTGTC GTCGCCGTCG GGCTCGTCGT CGCCGGCGAG	CGGCTTGTTC	TCGCCGCCGG CGGCGTCGGC GTGCTCGACG	GTGTGCTCGT GCCGATTCAA CGCGGCCTGC	GGAACCACGC	CGGTCACCGG TTTCCAGGTC GGTTACCGTC	GTCATGGTTG ATCTCCTGGC GTGGAATGTT	ACACCTTCCA CCACCACGAG AAGCTGCTAC	ອວອວວວອວວອ	ACCTGGCCAG CACCCGCCG AACCCGCGAG	CCCTTCGCCA TCAAGCTTTT TGGGCCCGCC
	CGGCCCATGG	CGCGGTGTTC	CGGCTCCACG	CGGTTGATGG	TCGCCGCCGG	GTGTGCTCGT	GGCGCGTCCG	CGGTCACCGG	GTCATGGTTG	ACACCTTCCA	CGTGAGAATC	ACCTGGCCAG	CCCTTCGCCA

10					;	57/63							
15 of 16	10980	11040	11100	11160	11220	11280	11340	11400	11460	11520	11580	11640	11700
FIGURE 3-15	GGCGAACCG ACCGTCAATC CGCGACGGCC GTCCCGCAAT	GTGACCACC GACCCGTCGG GCCAGGGTGA CATGGGCGGT	CCACTGACCG GGCAGGCTGT TGGCCATCGG CGCGGGCGCC AGGTGCGGGC CGCAGAGCCG	CAAAAGCTC GCTGGTCGGC ACCACCAGCC GGGTGAACAC	CACCGGCGC GCCGATCACG CAGTCCAGCG GCAGCCGACG	ATCGACCTC CGGGGCGATC CGTTCGGCCA CCGCCAGCGA	CGCCTGGCT GGGTATGCCG GCGGCGGCCA ACCCCGCCCA	GGTATCGCT GTCGAAGACC AGCTCGATCG AATGCACCAT	CAGTIGCGG TCGAACGCCG CCGCGCTCAT CGCCGCGAAG	GCCCCGGCG GGCAGAGCGG CCCGCACCGT AGCAATGCGC	GCTGCCACC AACCCGGGCG GGTCCGGCCA GCTGCCGATC	GTGCA GCAAGCGCTT CCAACGTCAA CTTGGTGTGG	CCACC ACCAAAGCCG CGGCGGCCAC GTCGACGGCG
	GCCGTCCCAG CGGCGCAAGC CGGCG1	GCGCAGCGCC CGCCCCAATT GGTGA(	3 GGCAGGCTGT TGGCC	GTGCACCTCG GCATGCAGGG CCAAA	GACATTGGCC CGCCCGAACA GCACCC	GGCAACCGCA CCCAGCGGCT CATCGA	CACGTGCGGA CGGCTGGCCG GCGCCT	GATGCGCCGG ATCGCCGCCT CGGTAI	CAGCCGACCA GCCCGGCAAC CCAGTI	TCCCCGGCAT CCAGCGACGC GGCCC	GCCAGCGCG ACCGATTCGA GGCTG(	ACCAGCCCTG CACATGAAAC CTGTTGTGCA GCAAGCGCTT	TTGAGGGTGC CCAGGTCCGC GGTGACCACC
	GCCGTCCCAC	GCGCAGCGC	CCACTGACCC	GTGCACCTCC	GACATTGGCC	GGCAACCGC	CACGTGCGG	GATGCGCCGC	CAGCCGACCA	TCCCCGGCAI	GCCAGCGCGC	ACCAGCCCTG	TTGAGGGTGC

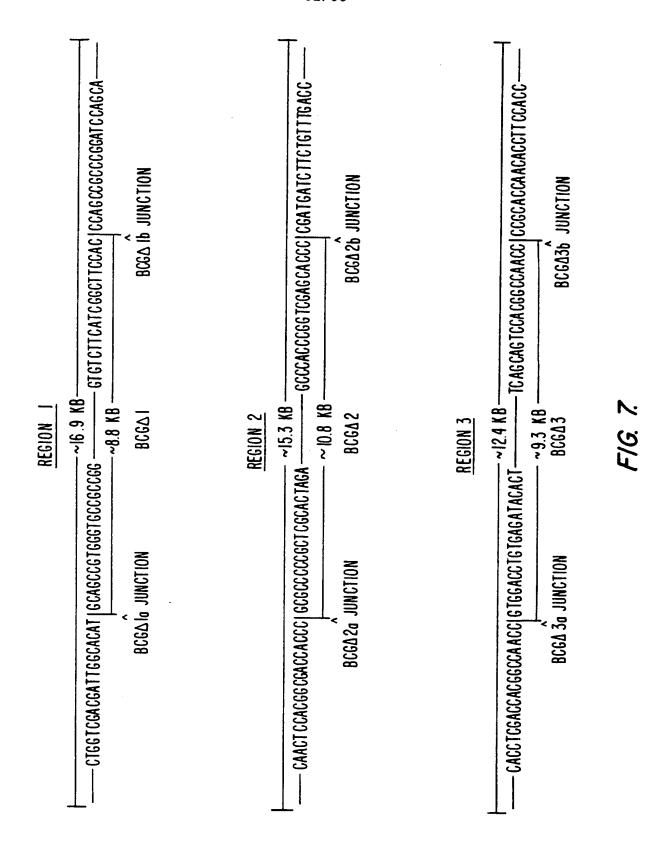
						58/	63					
16 of 16	11760	11820	11880	11940	12000	12060	12120	12180	12240	12300	12360	12412
FIGURE 3-16 Page	ACATCGCGCA GCGTGACGCC CGGCTCGGCG AGTTCGACCA GCAGCCCGCC CGCCCCTCG	ACGAGGGTCA ACCGCCCGGG ACGGTCCAGG TCTGCGATCA GCCGCACGAT CTGATCGCGG	GCGGGCAACG CCATCCCGGC GTGTTCGGCG GCGCCGCCG GGGCCATCGG CTGCGGATAT	CGCGCCAAGC CGGCCAGCTG GGTCACCCCG GCCAACCGGC CGACCTCGGC GAGGTCGTCG	TCACCGCGGG CGGTGCCGGT CTGAACGGGC TTGCACACCG CCACGTCGAT GCCGGCCTGA	CGTGCGGCCG ACGCCAGCGC CGCGCAGACG ACCGTCTTGC CGACCCCCGT GCCGGTCCCG	GIGACGACCA GGATCGICAA CGGCGCGCCA CGGCGAGAAC ATCCGICAGC ACCCGCCGGG	CCAGCTCGAG CTCGCCGGCG TTCAGCGATG CGCGCGCGGT CAGCCGCAGC CGCGACGTAC	CCGCGGGCAC CGTCGGCGGC CGGAAGCAGC CCACCTTGAC CCCGGCGTCC AGGCAGGCCG	CCGCGGCGGC CACTGCCGAC TCCGGCTCGC CCAGGATCAC CGACACCATC GCCGAGTCCG	GCACCGCAGC CACACCGCAC ATCCGCGCAA GTTCACCAGC GTGGTTGAGC ACCGCCTGCG	ATCGCCACGG CTCGGCCTGC AAGACGCGCCA GCGCGGCCCG TGCGGCACCT TC
	ACA	ACG	BOB	SBS	TCA	CGT	GTG	CCA	900	SOO	GCA	ATC

59/63 BCC∆ = ~ 8.8kb HOMOLOGUE ACCESSION 246257 U01072 X 79562 129506 ORF 2.3e-43 1.4e-14 3.0e-13 3.6e-16 ₩ 9 ≝ B. subtilis subtilisin SERINE PROTEASES M. tuberculosis esat6 M. leprae aceA BCG uraA £ 乭 MAX.~kDa) ENCODEO Protein 57 36 59 46 20 34 \$ **BINDING SITES** NONE GGA(11) AGGAGA (10) POSSIBLE RIBOSOME AGGA (10) 66A (4) 66A(9) GAGG (5) A66A(9) GAA (5) 66A(5) NONE =14823 - 13438 14643 - 13438 14541 - 13438 3130-4203 3139-4203 5075-6046 夂 13328-11946 2996-61901 16190 - 14820 889-2433 6954-8612 ೭ 1542 1368 ౼ 696 1386 102 1657 954 1380 ᅙ ORF | M. tuberculosis | CODON USAGE REGION 1 (16.9 kb) YES YES  $\leq$ YES YES YES YES YES YES 邑  $\leq$  $\underline{\infty}$ ఆ 9 ೨ 느  $\equiv$ 

F/G. 4.

							60/									
F/G. 5.		BCG ∆ = ~10.8 kb	ORF HOMOLOGUE ACCESSION #				P24194	U00015 A00975 U03393		X73226 X17445	A30545	X65104 U04851 Z22594 X59155				
<b>~</b> ≅-1∎⊥	- <del>ਨ</del>	908	P VALUE				9.9e -47 < le - 5	1.5e-7 ~4.e-5		9.9e - 146 2.7e - 36	6.7e-141	3.1e - 11 1.4e - 08 4.4e - 11 2.5e - 09				
Bg Kp Xb	11 12 13 14	26 21 21	HOMOLOGIES TO PREDICTED Encoded protein				E.coliciA lysR FAMILY	MLEPRAE COSMID BIG20 ORF Cutinases		Ecolisia Brouvwx	M. tuberculosis mpt 64	Ecol, gabP PERNEASE Styphimurium asp PERMEASE I harzianum indal gene RETROVIRAL RECEPTOR				
Bm Kp	- 0	<b>₹</b>	ENCODED PROTEIN (MAX.~kOd)	25	91	34	34	. 22	<u>6</u>	37	24		31	35	25	12
Bm Sp Bm	- 2	20 27	POSSIBLE Ribosome Binding sites	AGGGAG (7)	AGAA (4)	NONE	AG? (8) GGA (8)	NONE	NONE	AGGA (11)	AAGA (6)	AG (10)	GGAAGA (6)	GAG (10) GGAA (8) GGA (9)	NONE	AG (10)
H3 Bm	4 5 6	21 20	TOP IIRS)	1829-2386	2862-3298	3003-3590	5187-6134 5376-6134	6561-7217	8036-8560	9941-10909	11118-11783	11965-13407	14221-13376	8259-7211 7939-7211 7931-7211	4992-4327	5117-4521
<b>3</b>	2 3	2A 282	281 ORF SIZE (BASE PAIRS)	829	437	588	948	657	522	996	999	1443	846	1050	999	597
REGION 2 (15.3 KB)  Bm xb Bg	0		M. tuberculosis CODON USAGE	YES	YES	YES	YES	YES	YES		YES	YES	YES	YES	<b>S</b> ES	YES
뛽			ORF	24	281	282	32	20	2E	2F	23	2H	2	1.2	2K	<b>%</b>

		<del></del> -		61/6	3	<del></del>	· <del>-</del>		· · · ·						_
<b>6</b>	9.5 KB	HONOLOGUE Accession #	X70901 U00021	090361 M29040 K00676 X01805, X07724		X76288	X76288 L 37531				0100010	000000	N29292	0100010	
F/G 6.	BUGA = ~3.3 KB	P VALUE	2.9e-64 5.1 e-13	3.0e-05 7.8e-4 8.2e-4		4.2e -26	3.2e - 11				6.2e-69	6.9e-53	2. be - Uo 1.4e - O5	1.0e-81	
Ps Bg H3 R5 Xh	31	HONOLOGIES TO PREDICTED ENCODED PROTEIN	MTB mce sau3A M. leprae cosmid L247	ACTINOPHAGE R4 011P GENE A subfit site spec. Recomb. Recombinases/invertases		<i>S. coelicolor</i> phage phi-G3I EARLY REGION	paly palz GENES				M. leprae B1170	M. leprae bioDAYB	<i>c. gluramicum</i> biou <i>B. sphaericus</i> bio DAYB	M. Jeprae cosmid BII70	
2 S - 1 - 1 - 1 - 1 - 1 - 1 - 1 - 1 - 1 -		ENCODED PROTEIN (MAXkDa)	45	49	61	34	21	20	19	47	28	1	30	25	22
Kp Ps RI		POSSIBLE RIBOSOME BINDING SITES	NONE		66A (6)	NONE GGA (7)	NONE GGAA(5)	GA (9)	GAAGG (8)	GA (8)	GGA A G (6)		GAA (II)	AGG (4)	AG (10)
Kp Ri	36 3E	START - STOP (BASE PAIRS)	613-1755	1214-2560	2820 - 3332	4007 - 4930 4070 - 4930	4795-5337 4915-5337	5639-6214	5253-7762 7285-7762	7868 - 9197	10146 - 11810 10164 - 11810		12319-11402	11594-10893	10147-779488
~	38	ORF SIZE (BASE PAIRS)	1143	1347	513	924	543	576	210	1330	9991		816	702	099
REGION 3 (12.4 KB)  RI Kp	3A	M. tuberculosis CODON USAGE	YES	YES	YFS	98 8	NO YES	YES	YES	YES	YES		YES	YES	YES
		ORF	3A	38	35	8	35	7.	36	₩.	31		<del></del>	3	2



SUBSTITUTE SHEET (RULE 26)

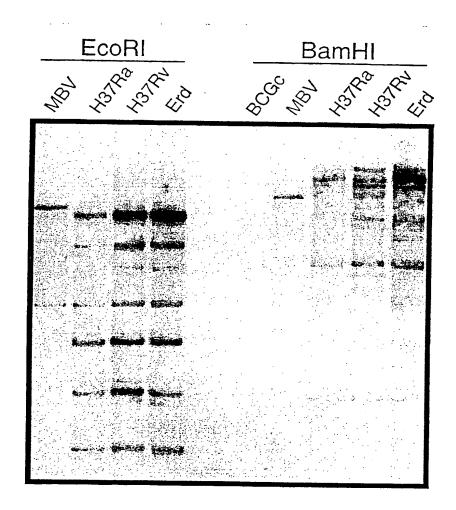


Figure 8

International application No. PCT/US96/01938

h national classification and IPC				
ed by classification symbols)				
	5.5, 24.32, 24.33			
ne extent that such documents are inc	luded in the fields searched			
	· · · · · · · · · · · · · · · · · · ·			
name of data base and, where practic	cable, search terms used)			
ppropriate, of the relevant passages	Relevant to claim No.			
61, No. 5, issued M	ay 1-10, 16, 17,			
sence of the MPB64 ge	ne 24, 25			
erium bovis BCG", pag	1			
	18-23			
NI	00 1 7			
AL/ OZ OCTOBEL 1989, S	ee   1-7			
9, No. 10, issued Octob	per 1-17			
on, characterization a	nd			
ycobacterium tuberculo:	sis			
pecies-specific : rauence	ε",			
ument.				
See patent family anne	х.			
"T" later document published after the	he international filing date or priority			
principle or theory underlying the	he invention			
"X" document of particular relevan considered novel or cannot be co	ce; the claimed invention cannot be			
considered to involve an invo	entive step when the document is			
being obvious to a person skille	er such documents, such combination d in the art			
*&* document member of the same	patent family			
Date of mailing of the international	al search report			
29 MAY 1996				
Authorized officer	A FAMAS TAN			
William In 4/000.				
I JULIAL FREDMAN				
	'Y' document of particular relevant considered to involve an invombined with order of the order			

Form PCT/ISA/210 (second sheet)(July 1992)\*

International application No. PCT/US96/01938

_		T
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	Abstracts of the 1994 IDSA Annual Meeting, Clin. Infect. Dis., Volume 19, issued October 1994, R. Frothingham et al, "Sequence based strain differentiation in the Mycobacterium tuberculosis complex, including rapid identification of M. bovis BCG", page 565, see abstract 10.	1-25
X	R. GHERNA et al, "AMERICAN TYPE CULTURE COLLECTION: CATALOGUE OF BACTERIA AND PHAGES", Eighteenth edition, published 1992, pages 202 and 211, see entire document.	11-15
X	Infection and Immunity, Volume 62, No. 4, issued April 1994, L.	1-7, 16-25
Y	Pascopella et al, "Use of in vivo complementation in Mycobacterium tuberculosis to identify a genomic fragment associated with virulence", pages 1313-1319, see entire document.	26
Y	Science, Volume 261, issued 10 September 1993, S. Arruda et al, "Cloning of an M. Tuberculosis DNA fragment associated with entry and survival inside cells", pages 1454-1457, see entire document.	1-23
X	US,A,5,171,839 (PATARROYO) 15 December 1992, columns 5-	1-10
Y	10.	16-23
Y	Nature, Volume 256, issued 07 August 1975, C. Kohler et al, "Continuous cultures of fused cells secreting antibody of predefined specificity", pages 495-497, see entire document.	10
Y	US,A, 4,683,202 (MULLIS) 28 July 1987, see entire document.	16-22, 24, 25
Y	Genomics, Volume 4, issued 1989, D. Wu et al, "The ligation amplification reaction (LAR) amplification of specific DNA sequences using sequential rounds of template directed ligation", pages 560-569, see figure 2.	16-22, 24, 25
Y	US,A, 4,410,660 (STRAUS) 18 October 1983, columns 14 and 15.	23
Y	Gene, Volume 131, issued 1993, A. Kinger et al, "Identification and cloning of genes differentially expressed in the virulent strain of mycobacterium tuberculosis", pages 113-117, see page 114, column 2.	1-26
X,P	WO,A2,95/17511 (JACOBS ET AL) 29 June 1995, see entire	1-26

International application No. PCT/US96/01938

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No
ζ,E	J. Bacteriol., Volume 178, No. 5, issued March 1996, G. Mahairas et al, "Molecular analysis of genetic differences between mycobacterium bovis BCG and virulent M. bovis", pages 1274-1282, see entire document.	1-26
/, P	Microbiology, Volume 141, issued 1995, J. Rodriguez et al, "Species-specific identification of mycobacterium bovis by PCR", pages 2131-2138, see entire document.	1-7, 16-22, 24, 25
۲	Hybridoma, Volume 13, No. 1, issued 1994, A. Arya et al,	8-10
·· ?	"Production and characterization of new murine monoclonal antibodies reactive to mycobacterium tuberculosis", pages 21-30, see page 27, table 1.	16-23

International application No. PCT/US96/01938

A. CLASSIFICATION OF SUBJECT MATTER:

IPC (6):

C12Q 1/68; G01N 33/53; C12P 19/34; C12N 5/10, 1/21; C07K 5/00, 14/00, 16/00; C07H 21/02, 21/04

A. CLASSIFICATION OF SUBJECT MATTER:

US CL :

435/6, 7.1, 91.1, 91.2, 240.2, 252.3; 530/300, 350, 387.1, 388.1; 536/22.1, 23.1, 24.3, 24.32, 24.33

#### **B. FIELDS SEARCHED**

Electronic data bases consulted (Name of data base and where practicable terms used):

### APS, MEDLINE, BIOSIS, CAPLUS, WPIDS

search terms: mycobacter?, tubercul?, bovis?, BCG, calmette, guerin, DNA, RNA, oligo, nucleic, oligonucleotide, hybridi?, probe, primer, amplif?, PCR, polymerase chain, ligase chain, LCR, attenuat?, immunoassay, antibod?, monoclon?, polyclon?, protein, peptide, antigen, virulenc?, infect?

Form PCT/ISA/210 (extra sheet)(July 1992)\*

THIS PAGE BLANK (USPTO)

Value of Proce

```
standard; DNA; PRO; 1069 BP.
     AF004671
IĎ
XX
     AF004671;
4G
XX
     q3253155
NI
XX
     29-JUN-1998 (Rel. 56, Created)
ЭT
     29-JUN-1998 (Rel. 56, Last updated, Version 1)
DΤ
XX
     Mycobacterium tuberculosis H37Rv esat6 promoter region, L45 antigen
DE
     homologous protein LHP (lhp) gene, complete cds, and early secreted antigenic target 6 kDa (esat6) gene, partial cds.
DΕ
DΕ
XX
KW
XX
     Mycobacterium tuberculosis
OS
     Eubacteria; Firmicutes; Actinomycetes; Mycobacteria; Mycobacteriaceae;
OC
     Mycobacterium.
OC
XX
     [1]
RN
     1-1069
RP
     Berthet F.-X., Birk Rasmussen P., Andersen P., Gicquel B.;
RA
     "Promoter analysis of the M. tuberculosis orf1C gene encoding the early
RT
     secreted antigenic target 6 kDa (ESAT-6)";
RT
     Unpublished.
RL
XX
RN
      [2]
      1-1069
RP
     Berthet F.-X., Birk Rasmussen P., Andersen P., Gicquel B.;
RA
RT
     Submitted (19-MAY-1997) to the EMBL/GenBank/DDBJ databases.
RL
     Mycobacterial Genetics Unit, Institut Pasteur, 25, rue de Dr Roux,
RL
     Paris, 75 75724, France
RL
XX
                       Location/Qualifiers
FH
      Key
FH
                       1. .1069
FT
      source
                       /organism="Mycobacterium tuberculosis"
FT
                       /chromosome="Region of difference RD1"
FT
                       /strain="H37Rv"
FT
                       525. .827
      CDS
FT
                       /codon_start=1
FT
                       /db_xref="PID:g3253156"
FT
                       /note="culture filtrate protein 10 kDa CFP-10"
 FT
                       /transl_table=11
 FT
                       /gene="lhp"
 FT
                       /product="L45 antigen homologous protein LHP"
 FT
                       /translation="MAEMKTDAATLAQEAGNFERISGDLKTQIDQVESTAGSLQGQWR
 FT
                       AAGTAAQAAVVRFQEAANKQKQELDEISTNIRQAGVQYSRADEEQQQALSSQMGF"
 FT
                       860. .>1069
 FT
      CDS
                       /codon_start=1
 FT
                       /db_xref="PID:g3253157"
 FT
                       /note="secreted T cell antigen; ESAT-6"
 FT
                       /transl_table=11
 FT
                       /gene="esat6"
 FT
```

THIS PAGE BLANK (USPTO)

BESTFIT of: Sa202820\_0001.Dna check: 1179 from: 1 to: 1277 (i) APPLICANTS: (A) NAME: INSTITUT PASTEUR (A) NAME: STATENS SERUM INSTITUT (ii) TITLE OF INVENTION: A POLYNUCLEOTIDE FUNCTIONALLY CODING FOR THE LHP PROTEIN FROM MYCOBACTERIUM TUBERCULOSIS, ITS BIOLOGICALLY ACTIVE DERIVATIVE FRAGMENTS, AS WELL AS from: 1 to: 1069 check: 4782 to: R55u027.Af004671 AF004671 standard; DNA; PRO; 1069 BP. ID XXAF004671; ACXX q3253155 NIXX Symbol comparison table: Gencoredisk:[Gcgcore.Data.Rundata]Swgapdna.Cmp CompCheck: 2335 Average Match: 10.000 50 Gap Weight: Average Mismatch: -9.000 3 Length Weight: 1069 Length: 10595 Ouality: Gaps: 9.911 Ratio: Percent Identity: 99.532 Percent Similarity: 99.532 Match display thresholds for the alignment(s): = IDENTITY 5 1 Sa202820\_0001.Dna x R55u027.Af004671 October 21, 1998 14:36 ... 1 CTGCAGCAGGTGACGTCGTTGTTCAGCCAGGTGGGCGGCACCGGCGGCGG 50 1 ĊŦĠĊĀĠĊĀĠĠŦĠĀĊĠŦĊĠŦŦĠŦŦĊĀĠĊĊĀĠĠŦĠĠĠĊĠĠĊĀĊĊĠĠĊĠĠĊĠĠ 51 CAACCCAGCCGACGAAGCCGCGCAGATGGGCCTGCTCGGCACCAGTC 100 51 CAACCCAGCCGACGAAGCCCGCAGATGGGCCTGCTCGGCACCAGTC 100 151 GGCCTGCTGCGCGGGGTCGCTACCTGGCGCAGGTGGGTCGTTGACCCG 200

THIS PAGE BLANK (USPTO)

201 CACGCCGCTGATGTCTCAGCTGATCGAAAAGCCGGTTGCCCCCTCGGTGA 250
201 CACGCCGCTGATGTCTCAGCTGATCGAAAAGCCGGTTGTGTGTG
251 TGCCGGCGGCTGTTGCCGGATCGTCGGTGACGGGTGGCGCCGCTCCGGTG 300
301 GGTCCGGGAGCGATGGGCCAGGGTTCGCAATCCGGCGGCTCCACCAGCCC 350
and agamentage coccoccoccoccoccoccoccoccoccoccoccoccoc
351 GGGTCTGGTCGCGCCGGCACCGCTCGCGCAGGAGCGTGAAGAAGACGACG 400 351 GGGTCTGGTCGCGCGCGCACCGCTCGCGCAGGAGCGTGAAGAAGACGACG 400
401 AGGACGACTGGGACGAAGAGGACGACTGGTGAGCTCCCGTAATGACAACA 450
401 AGGACGACTGGGACGACGACTGGTGAGCTCCCGTATTGATGTGAGCTCCCGTATTGATGTGAGCTCCCGTATTGATGTGAGCTCCCGTATTGATGTGAGCTCCCGTATTGATGTGAGCTCCCGTATTGATGTGAGCTCCCGTATTGATGTGAGCTCCCGTATTGATGTGAGCTCCCGTATTGATGATGTGAGCTCCCGTATTGATGTGAGCTCCCGTATTGATGTGAGCTCCCGTATTGATGTGAGCTCCCGTATTGATGTGAGCTCCCGTATTGATGTGAGCTCCCGTATTGATGTGAGCTCCCGTATTGATGTGAGCTCCCGTATTGATGTGAGCTCCCGTATTGATGTGAGCTCCCGTATTGATGTGAGCTCCCGGTATTGATGTATTGATGTGAGCTCCCGGTATTGATGTATTGATGTATGATGTATTGATGA
451 GACTTCCCGGCCACCCGGGCCGGAAGACTTGCCAACATTTTGGCGAGGAA 500
501 GGTAAAGAGAAAGTAGTCCAGCATGGCAGAGATGAAGACCGATGCCGC 550
501 GGTAAAGAGAAAGTAGTCCAGCATGGCAGAGATGAAGACCGATGCCGC 550
551 TACCCTCGGGCAGGAGGCAGGTAATTTCGAGCGGATCTCCGGCGACCTGA 600
551 TACCCTCGCGCAGGAGGCAGGTAATTTCGAGCGGATCTCCGGCGACCTGA 600
601 AAACCCAGATCGACCAGGTGGAGTCGACGGCAGGTTCGTTGCAGGGCCAG 650
651 TGGCGCGCGGCGGGGGCCGCCCAGGCCGCGGTGGTGCGCTTCCA 700
651 TGGCGCGCGCGGGGGACGGCCGCCCAGGCCGCGGTGGTGCGCTTCCA 700
701 AGAAGCAGCCAATAAGCAGAAGCAGGAACTCGACGAGATCTCGACGAATA 750
701 AGAAGCAGCAATAAGCAGAAGCAGGAACTCGACGAGATA 750
751 TTCGTCAGGCCGGCGTCCAATACTCGAGGGCCGACGAGGAGCAGCAGCAG 800 
801 GCGCTGTCCTCGCAAATGGGCTTCTGACCCGCTAATACGAAAAGAAACGG 850

THIS PAGE BLANK (USPTO)

851	AGCAAAAACATGACAGAGCAGTGGAATTTCGCGGGTATCC	900
		900
901		950
901	GGCAAGCGCAATCCAGGGAAATGTCACGTCCATTCATTCCCTCCTTGACG	950
951	AGGGGAAGCAGTCCCTGACCAAGCTCGCAGCGGCCTGGGGCCGTAGCGGT	1000
951		1000
	TCGGAGGCGTACCAGGGTGTCCAGCAAAAATGGGACGCCACGGCTACCGA	1050
1001		
1001	TCGGAGGCGTACCAGGTGTCCAGCAAAAATGGGACGCCACGGCTACCGA	1050
1051	GCTGAACAACGCGCTGCAG 1069	
1051		

.

THIS PAGE BLANK (USPTO)